

OM of: US-09-236-995D-2 to: GenEmbl.* out_format : pfs
Date: Mar 9, 2002 7:40 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlp
-Q/cgnd_1/USPTO.spool/US09236995/runat_08032002_113149_25104/app_query.fasta_1.1273
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.100 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09236995 -CGN1_1_8116 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-236-995D-2
Query length: 982
Database: GenEmbl.*
Database sequences: 1472140
Database length: 341344837
Search time (sec): 3984.260000

| score_list: | Strd Orig | ZScore | Escore | Len | Documentation |
|-----------------|-----------|---------|----------|--------|---|
| gb_pl:AF093627 | + 4742.00 | 6447.72 | 0.0 | 3285 | ! AF093627 Zea mays poly(ADP-ribose) |
| gb_pl:ZMPARP1 | + 4666.50 | 6344.95 | 0.0 | 3211 | ! AJ222589 Zea mays mRNA for poly(ADP-ribose) |
| gb_pl:ATH131705 | + 2774.50 | 3764.21 | 2.2e-201 | 3187 | ! AJ131705 Arabidopsis thaliana |
| gb_pl:AC006593 | - 1973.50 | 2640.95 | 7.9e-139 | 79663 | ! AC006593 Arabidopsis thaliana |
| gb_ro:AF168781 | + 1602.50 | 2165.95 | 2.3e-112 | 3047 | ! AF168781 Arabidopsis thaliana |
| gb_pat:AS2134 | + 1599.50 | 2159.77 | 5.0e-112 | 3792 | ! AS2134 Sequence 1 from Patent |
| gb_ro:HUMPOLP | + 1597.00 | 2156.75 | 7.4e-112 | 3640 | ! M18112 Human poly(ADP-ribose) |
| gb_ov:GGADPRP | + 1594.50 | 2155.07 | 9.2e-112 | 3036 | ! Y52690 Chicken mRNA for poly(ADP-ribose) |
| gb_ro:HUMR12041 | + 1593.00 | 2150.90 | 1.6e-111 | 3795 | ! X03473 Human poly(ADP-ribose) |
| gb_ro:BC012041 | + 1593.00 | 2150.78 | 1.6e-111 | 3845 | ! BC012041 Mus musculus, Simlana |
| gb_ro:RN094340 | + 1592.00 | 2151.38 | 1.5e-111 | 3128 | ! U94340 Rattus norvegicus poly(ADP-ribose) |
| gb_pat:AL14359 | + 1587.00 | 2142.84 | 4.4e-111 | 3747 | ! I14359 Sequence 5 from patent |
| gb_ro:MMADPRP | + 1584.50 | 2141.01 | 5.6e-111 | 3172 | ! I14206 Mouse mRNA for poly(ADP-ribose) |
| gb_pat:AX058340 | + 1584.00 | 2140.72 | 5.8e-111 | 3045 | ! AX058340 Sequence 24 from Patent |
| gb_pat:AX062277 | + 1584.00 | 2140.72 | 5.8e-111 | 3045 | ! AX062277 Sequence 136 from Patent |
| gb_ro:HUMPPOL | + 1584.00 | 2138.97 | 7.3e-111 | 3660 | ! M32721 Human poly(ADP-ribose) |
| gb_ov:HELADPRP | + 1576.00 | 2134.73 | 1.3e-110 | 3718 | ! D14567 Frog mRNA for poly(ADP-ribose) |
| gb_ov:BOVPAS | + 1576.00 | 2128.24 | 2.9e-110 | 3589 | ! D90073 Bovine poly(ADP-ribose) |
| gb_ov:XLPARP | + 1548.00 | 2089.98 | 3.9e-108 | 3617 | ! Z12139 X. laevis PARP gene encoding |
| gb_pat:AX058362 | + 1518.50 | 2050.90 | 5.8e-106 | 3200 | ! AX058362 Sequence 46 from Patent |
| gb_in:SPEDPRP | + 1381.00 | 1862.59 | 1.8e-95 | 3463 | ! D1482 Sarcophaga peregrina mRNA |
| gb_ro:AL442128 | + 1285.50 | 1696.13 | 3.4e-86 | 155026 | ! AL442128 Human DNA sequence |
| gb_in:ROADPRP | + 1269.50 | 1711.61 | 4.6e-87 | 3080 | ! D13806 Fruit fly mRNA for poly(ADP-ribose) |
| gb_ro:CORADPRP | + 1243.50 | 1670.14 | 9.4e-85 | 5787 | ! L24094 Gorilla gorilla ADP-ribose |
| gb_pat:IL14360 | + 1237.00 | 1662.03 | 2.7e-84 | 9254 | ! I14360 Sequence 7 from patent |
| gb_ro:HUMADPRP2 | + 102.00 | 1351.99 | 5.0e-67 | 1771 | ! M17081 Human placental poly(ADP-ribose) |
| gb_ro:AF126717 | + 993.50 | 1340.61 | 2.1e-66 | 1732 | ! AF126717 Mus musculus mPARP mRNA |
| gb_ro:RNPARP2 | + 990.50 | 1335.44 | 4.1e-66 | 1939 | ! X65497 R. norvegicus mRNA for poly(ADP-ribose) |
| gb_pat:IL14357 | + 948.50 | 1317.35 | 4.2e-65 | 2682 | ! I14357 Sequence 3 from patent |
| gb_pat:AX062218 | + 898.00 | 1204.18 | 8.5e-59 | 3308 | ! AX062218 Sequence 177 from Patent |
| gb_ro:HS236912 | + 851.50 | 1146.79 | 1.3e-55 | 1754 | ! AJ236912 Homo sapiens mRNA for poly(ADP-ribose) |
| gb_ro:HS236876 | + 851.50 | 1146.56 | 1.4e-55 | 1797 | ! AJ236876 Homo sapiens mRNA for poly(ADP-ribose) |
| gb_pat:AX058317 | + 851.50 | 1146.47 | 1.4e-55 | 1814 | ! AX058317 Sequence 1 from Patent |
| gb_pat:AX058317 | + 851.50 | 1146.32 | 1.4e-55 | 1843 | ! AX058317 Sequence 1 from Patent |
| gb_pat:AX058374 | + 851.50 | 1146.16 | 1.4e-55 | 1874 | ! AX058374 Sequence 58 from Patent |
| gb_pat:AX058374 | + 851.50 | 1144.87 | 1.7e-55 | 2147 | ! Z48243 A. thaliana PARP mRNA for poly(ADP-ribose) |
| gb_ro:ATPARP | + 842.50 | 1133.71 | 7.2e-55 | 1910 | ! AF085734 Homo sapiens NAD+ ADP-ribose |
| gb_pat:AX058338 | + 840.00 | 1131.36 | 9.7e-55 | 1707 | ! AX058338 Sequence 22 from Patent |
| gb_ro:MMU7780 | + 840.00 | 1131.36 | 9.7e-55 | 1707 | ! AJ007780 Mus musculus mRNA for poly(ADP-ribose) |

gb_ro:AF072521 + 840.00 1131.04 1.0e-54 1766 ! AF072521 Mus musculus poly(ADP-ribose)
gb_pl:ZMPARP1 + 838.50 1126.50 1.8e-54 2295 ! AJ222588 Zea mays mRNA for poly(ADP-ribose)
gb_pr:AK001980 + 838.00 1127.22 1.6e-54 1980 ! AK001980 Homo sapiens cDNA
gb_in:DMPARP5 + 830.50 1117.79 5.5e-54 1822 ! AF051548 Drosophila melanogaster

seq_name: gb_pl:AF093627

seq_documentation_block:
LOCUS AF093627 3285 bp mRNA PLN 29-NOV-1998
DEFINITION Zea mays poly(ADP-ribose) polymerase (PARP1) mRNA, complete cds.
ACCESSION AF093627
VERSION AF093627.1 GI:3928870
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 3285)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
AUTHORS Mahajan, P.B. and Zuo, Z.
TITLE Purification and cDNA cloning of maize Poly(ADP-ribose) polymerase
JOURNAL Plant Physiol. 118 (3), 895-905 (1998)
MEDLINE 99026291

REFERENCE 2 (bases 1 to 3285)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
AUTHORS Mahajan, P.B. and Zuo, Z.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1998) Crop Protection, Pioneer HiBred International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131-1004, USA

FEATURES
source

Location/Qualifiers
1..3285
/organism="Zea mays"
/db_xref="taxon:4577"
1..3285
/gene="PARP1"
100..3042
/gene="PARP1"
/EC_number="2.4.2.30"
/function="add ADP-ribose"
/codon_start=1
/product="poly(ADP-ribose) polymerase"
/protein_id="AAC79704.1"
/db_xref="GI:3928871"
/translation="MAAPPKAAEAKSASCKSRSPFIADQLRLGKMWQASQFD
GFPMNHAICFISKNQIKSVDDVDGIDLRWDDQKIRNYVGSASGTSSTAAPE
KCTIETAPARTSCRCSEKRTIGSVRLSAKLESGKPGIPWTHANCFEVSFAPE
KFGWDTLSDDEKRTLDLVKQDVGNEQNKGRKKSENDIDSYKARLDEANGQDT
VRNGQLVDPGRGNTSADQLKQSDTLWKLDKGLTHVSAELRDLMEANGQDT
SGPERHLIDRCADGMLFGALGPCVANGMYVNGOYCGSNVSEKCTYSATEPVR
VKKWQIPHTKNDYLMKWKSKQKPKPVLPPMSPEKSGSKATORTSILSSKGLDK
LRFVSQGSKEANEWEKLEKLAGANFYARVKDIDCLACGLDNEAENAEVRARLK
IPIVREGYIGECVKNKKNKPFLLYLENALSQSTVTIVKVGSAVHESGLQDTA
HILEDGKSIYATNLNSDLALGVNSYVLIQIEQDQSGECYVERKVGWVGSEKIGSQK
LEMSKTEALKEKRLFLKGTGNSWEAKCTNFKPKQPFYPLDVIDYGVKKAPRKRD
ISEMKSALPQLLEKMLPNVETIRAAAMFEFINMSEPLKSLKENTKGFPEALTE
IONLKTADAQALAVRESLIVAAASNFETLIPSHIPI INDEDDLMKAKMLBALQDI
EASKIVGDSDESDDDKMKMLHCDDITPLANDSEYDLKIEYLINHTAPHTKMSL
ELEEFSLDSDGELNRYKNNLNHNKMLGSLRTNFVGLISQSLGRLTAPPEPVYG
YMGKGLGYFADLVSKAQYCYVDNRNPNVGLMLSEVALGDMYELKATSMKPPKGRK
STKGLGKTVPLESEFVKVRDDVVVPCGKVPSSRSSELMEYINVTYSQVRKMQFL
KVRPHKR"

BASE COUNT 1003 a 617 c 828 g 837 t
ORIGIN

alignment_scores:
Quality: 4742.00 Length: 985
Ratio: 5.002 Gaps: 5
Percent Similarity: 96.244 Percent Identity: 95.635

alignment_block:
US-09-236-995D-2 x AF093627

Align seg 1/1 to: AF093627 from: 1 to: 3285

```

1 MetAlaAlaProProLysAlaTrpLysAlaGluTyrAlaLysSerGlyAr 17
|||||
100 ATGCGCGCGCGCCAAAGGCGTGAAGCGGAGTATGCCAAGTCTGGGCG 149
|||||
17 gAlaSerCysLysSerCysArgSerProIleAlaLysAspGlnLeuArgL 34
|||||
150 GGCTCTGTCGCAAGTCATGCGCGTCCCTATCGCCAAGGACCAAGTCCGTC 199
|||||
34 euGlyLysMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrp 50
|||||
200 TTGGCAAGATGTTACAGCGCTCACAGTTCGACGGCTTCATGCCGATGG 249
|||||
51 AsnHisAlaArgCysIlePheSerLysLysAsnGlnIleLysSerValAs 67
|||||
250 AACCATGCCAGGTGCATCTTCAGCAAGAGAACCAAGATAAATCCGTGA 299
|||||
67 pAspValGluGlyIleAspAlaLeuArgTrpAspAspGlnGluLysIleA 84
|||||
300 CGATGTTGAAGGGATAGATGCATCTTAGATGGGATGATCAAGAGAAGATAC 349
|||||
84 rqAsnTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaPro 100
|||||
350 GAACTACCTCGGAGTGCCTCAGCTGGTACAAAGTTCACAGCTGCTCCT 399
|||||
101 ProGluLysCysThrIleGluIleAlaProSerAlaArgThrSerCysAr 117
|||||
400 CCTGGAATGTACAAATTGAGATTGCTCCATCTGCCCGCTACTTCATGTAG 449
|||||
117 qArgCysSerGluLysIleThrLysGlySerValArgLeuSerAlaLysL 134
|||||
450 ACGATGCAAGTGAAGAGATTACAAAAGGATCGGTCCGCTCTTCAGCTAAC 499
|||||
134 euGluSerGluGlyProLysGlyIleProTyrHisAlaAsnCysPhe 150
|||||
500 TTGACAGTGAAGTCCCAAGGATATACATGGTATCATGCCACTGTTTC 549
|||||
151 PheGluValSerProSerAlaThrValGluLysPheSerGlyTrpAspTh 167
|||||
550 TTTGAGGATATCCCGCTCTGCAACTGTGTGAGAAGTTCACAGCTGGGATAC 599
|||||
167 rIeuSerAspGluAspLysArgThrMetLeuAspLeuValLysLysaspv 184
|||||
600 TTTGCCATGAGGATAGAGAACCATGCTCGATCTTGTAAAAAAGATG 649
|||||
184 alGlyAsnAsnGluGlnAsnLysGlySerLysArgLysLysSerGluAsn 200
|||||
650 TTGGCAACAATGAACAAAATAAGGTTCCCAAGCCCAAGAAAGTGAAT 699
|||||
201 AspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrSerGluGl 217
|||||
700 GATATTGATAGCTACAAATCCGCCAGGTAGATGAAAGTACATCTGAAG 749
|||||
217 yThrValArgAsnLysGlyGlnLeuValAspProArgGlySerAsnThrS 234
|||||
750 TACACTCGGAACAAAGGGCAACTTTAGACCCACGTTGGTTCCAATACTA 799
|||||
234 erSerAlaAspIleGlnLeuLysLeuLysGluGlnSerAspThrLeuTrp 250
|||||
800 GTTCAGCTGATATCCAACCTAAAGCTTAAGGAGCAAAAGTGACACACTTTGG 849
|||||
251 LysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGluLeuArgAs 267
|||||
850 AAGTTAAGGATGGACTTAAGACTCATGTATCGGCTGCTGAATTAAGGGA 899
|||||
267 pMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArgHisLeuL 284
|||||
900 TATCTTGGAGCTAAATGGCAGGATACATCAGGACCAAGAACGACCTAT 949
|||||
284 euAspArgCysAlaAspGlyMetLeuPheGlyAlaLeuGlyProCysPro 300
|||||
950 TGGATCGCTGGCGGATGGAATGCTATTTGGAGCGCTGGGCTCCTTGCCCA 999

```

```

301 ValCysAlaAsnGlyMetTyrTyrTyrAsnGlyGlnTyrGlnCysSerGl 317
|||||
1000 GTCTGTGCTAATGCGATGTACTATTATAAGTGCAGTACCAATGCACTGG 1049
|||||
317 yAsnValSerGluTrpSerLysCysThrTyrSerAlaThrGluProValA 334
|||||
1050 TAAATGTGTGAGAGTGGTCCAAAGTGTACATATCTGTGCCACAGAACCTGTCC 1099
|||||
334 rgValLysLysLysTrpGlnIleProHisGlyThrLysAsnAspTyrLeu 350
|||||
1100 GCGTTAAAGAAAGTGGCAAAATCCACATGCACAAAGAAATGATTACCTT 1149
|||||
351 MetLysTrpPheLysSerGlnLysValLysLysProGluArgValLeuPr 367
|||||
1150 ATGAAGTGGTTCAAAATCTCAAAAGGTTAAGAAACCCAGAGAGGGTTCCTTC 1199
|||||
367 oProMetSerProGluLysSerGlySerLysAlaThrGlnArgThrSerL 384
|||||
1200 ACCAATGTCACCTGAGAAATCTGGAAGTAAAGCAACTCAGAGAACATCAT 1249
|||||
384 euLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValGlyGln 400
|||||
1250 TGCTGTCTTCTAAAGGGTGGATAAAATTAAGGTTTCTGTGTAGGACAA 1299
|||||
401 SerLysGluAlaAlaAsnGluTyrIleGluLysLeuLysLeuAlaGlyAl 417
|||||
1300 TCAAAAGAACACCAATAGTGGATGAGAACTCAGAACTCAAACTTCTGTGTG 1349
|||||
417 aAsnPheTyrAlaArgValValLysAspIleAspCysLeuIleAlaCysG 434
|||||
1350 CAACCTCTATGCCAGGTTGTCAAGATATTGATTGTTTAAATTGCATGTG 1399
|||||
434 lyGluLeuAspAsnGluAsnAlaGluValArgLysAlaArgArgLeuLys 450
|||||
1400 GTGAGCTCGACAATGAAATGCTGAAGTTCAGGAAGCAAGAGGAGGTGAAG 1449
|||||
451 IleProIleValArgGluGlyTyrIleGlyGluCysVal. LysArgThrL 467
|||||
1450 ATACCAATGTAGAGGAGGTTACATGGAGATGTGTTAAAAAGAACAA 1499
|||||
467 yScCysHisLeuIleCysIleAsn..TrpAsnAlaLeuGluSerSerL 483
|||||
1500 AATGCTGCCATTTGATTGTATAAACTAGAGAATGCTTAGAGTCTCTCAA 1549
|||||
483 ySgLy***ThrValThrValLysValLysGlyArgSerAlaCysSer*** 499
|||||
1550 AAGGCAGTACTGTCACTGTAAAGTTAAGGGCCGAAGTCTGTTTCATGAG 1599
|||||
500 SerPro***ValCysLysAsnThrAlaHisIlePro****TrpGluLy 516
|||||
1600 TCCTCTGGTTTG...CAAGATACTGTCACATCTCTGA. AGATGGGAAA 1645
|||||
516 shIleGlnCys***LeuLysHisValLeuThr***His***ValCys. 532
|||||
1646 GCATATACAAATCAACCTTAAACATGCTGACCTGCG...ACTAGGTGTG 1692
|||||
533 ThrGlyTyrTyrValLeuGlnIleIleGluGlnAspAspGlySerGluCy 549
|||||
1693 AACAGCTACTATGTACTCCAGATCAATGAACAGGATGATGGTCTGAGTG 1742
|||||
549 sTyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGly 566
|||||
1743 CTACGTATTTCGTAAGTGGGCGGTTGGGAGTCAGAAAATTTGGAGGCG 1792
|||||
566 InLysLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysArg 582
|||||
1793 AAAAAGTGGAGAGATGCAAAACTGAGGCAATCAAGGAATTTCAAAAGA 1842
|||||
583 LeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysTh 599
|||||
1843 TTATTCTTGAGACACTGGAAACTCATGGGAGAGCTTGGGAATGTAAAC 1892

```

```

599 rAsnPheArgLysGlnProC
|||||
1893 CAATTTTCGAAGCAGCCT
|||||
616 lyValLysLysAlaProLy
|||||
1943 GTCTTAAGAAAGCACCAN
|||||
633 LeuAlaProGlnLeuLew
|||||
1993 CTTGCTCTCAATTGCTA
|||||
649 rTyrArgAlaAlaMetGluPheGlu***
|||||
2043 ATATAGAGCTGCTATGATGAATTTGAAATTA
|||||
666 euGlyLysLeuSer*****AsnIleGlu***Gly
|||||
2093 TTGGGAAGCTAAGCAAGAAATATTGAGAAAGGATTTGAAGCATTAACCT 2142
|||||
683 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu** 699
|||||
2143 GAGATACAGAAATTTATT.GAAGGACACCGCTGATCAAGCACTGGCTG.TT 2190
|||||
699 *GluLysAla**LeuLeu*****PheSerLeuLeuSerL 716
|||||
2191 AGAGAAAGCTTAATTGTGCTGCGAGCAATCGCTTTTCACTCTTATCCC 2240
|||||
716 euLeuPheIleLeuLeuLeuTyrGlyMetArgMetIleSerTyrSerLys 732
|||||
2241 TTCTATTATCATCTCATATTATACGGGATGAGGATGATTG.ATGATCAAA 2289
|||||
733 AlAlysMetLeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleVa 749
|||||
2290 GCGAAAATGCTTGAAGCTCTGCGAGGATATTGAAATTCITTCACAGATAGT 2339
|||||
749 lGlyPheAspSerAspSerAspGluSerLeuAspAspLysTyrMetIylsL 766
|||||
2340 TGGCTTCGATAGCGACAGTATGATGAATCTCTTGATGATAAATATATAAAC 2389
|||||
766 euHisCysAspIleThrProLeuAlaHisaspSerGluAspTyrLysLeu 782
|||||
2390 TTCACCTGACATCACCCGCTGGCTCAGCATAGTGAAGATTAACAAGTTA 2439
|||||
783 lIleGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSe 799
|||||
2440 ATTGAGCAGTATCTCCCAACACACATGCTCCTACTCACAGGACTGCTC 2489
|||||
799 rLeuGluLeuGluValPheSerLeuAspArgAspGlyGluLeuAsnL 816
|||||
2490 GCTGGAACCTGGAGCAAGTTTTTTCACCTGATCGAGATGGAGAACTTAATA 2539
|||||
816 ysTyrSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTyrPheHis 832
|||||
2540 AGTACTCAAGATATAAAAAAATATCTGCATACAACAGATGCTATTATGCGAC 2589
|||||
833 GlySerArgLeuThrAsnPheValGlyLysLeuSerGlnGlyLeuArgI 849
|||||
2590 GGTCAAGGTTGACGAATTTTGTGGGAATCTTCTACTCAAGGCTAAGAAT 2639
|||||
849 eAlaProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuT 866
|||||
2640 TGCACCTCTCAGGACCCCTGTTTACTGGCTATATGTCGCAAGGCTCT 2689
|||||
866 yrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArg 882
|||||
2690 ACTTTCAGATCTAGTAAGCAAGACCGCACAACTACTGTTATGCGATAGG 2739
|||||
883 AsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMe 899
|||||
2740 ATATATCTCTGATGGTTGATGCTTCTTCTGAGGTTGCTTTAGGAGACAT 2789
|||||
899 tTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysH 916

```

```

2790 GTATGAACTAAAGAACCCACGTCATGGACAAACCTCCAAGAGGGAAGC 2839
|||||
916 isSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 932
|||||
2840 ATTCGACCAAGGATTAGGCAAAACCGTGCCACTGGAGTACAGATTGTTG 2889
|||||
933 LysTrpArgAspAspValValProCysGlyLysProValProSerSe 949
|||||
2890 AAGTGGAGGATGATGCTAGTTCCCTGCGCAAGCGGTGCCATCATC 2939
|||||
949 rIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrS 966
|||||
2940 AATTAGGAGCTCTGAACATCATGTAATGAGTACATGCTTACACACAT 2989
|||||
966 erGlnValLysMetGlnPheLeuLeuLysValArgPheHisHisLysArg 982
|||||
2990 CCCAGGTGAAGTGCAGTCTTCTGCTGAAGTGCCTTTCCATCACAAAGG 3039
|||||

```

seq_name: gb_pl:ZMPARP2

seq_documentation_block:

LOCUS ZMPARP2 3211 bp mRNA PLN 19-NOV-1997

DEFINITION Zea mays mRNA for poly(ADP-ribose) polymerase (3211bp).

ACCESSION AJ222589

VERSION AJ222589.1 GI:2632128

KEYWORDS PARP gene; poly(ADP-ribose) polymerase.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 3211)
Babychuk, E., Cottrell, P., Storozhenko, S., Fuangthong, M.,
O'Farrell, M., Van Montagu, M., Inze, D. and Kushnir, S.

TITLE Higher plants possess two poly(ADP-ribose) polymerases

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3211)

AUTHORS Kushnir, S.

TITLE Direct Submission

JOURNAL Submitted (14-NOV-1997) Kushnir S., VIB, Dep.Genetics,

Ledeganckstraat 35, Gent, B9000, Belgium

FEATURES

source

/organism="Zea mays"

/db_xref="taxon:4577"

113..3022

/gene="pARP"

113..3022

/gene="pARP"

/function="secondary protein modification"

/codon_start=1

/product="poly(ADP-ribose) polymerase"

/protein_id="CAA10889.1"

/db_xref="GI:2632129"

/db_xref="SPTREMBL:O24570"

/translation="MAAPPAKAAEYAKSGRASCSCRSPIAKDOLGLKQWASOFD

GMPWNHNSVDDVEGIDALRWDDQEKIRNVSGASAGTSTAAPPKCIEIETAPSAR

TSCRRCEKIKTGSVRLSAKLESEGPKGIPIWHANCFFVSPSATVEKFSGWDTLSD

DKRTMLDLVKDDGNQNGSKRKKNDIDISYSARLDESTSEGTVRNKGQLVDR

GSNTSSADIQILKEQSDTLWLKGLKTHVSAALRDLMEANGODTSGPRHLLDR

ADGMIFGALGCPVCNMGVYNGYOCSGNVSEWSEKTSYATPEPVVKKKQIIPHGT

ANDYLMKFRSKQKVKPERVLPMSPKSGSKATQRTSLLSSKGLDIEIAKIVGQSK

AAANEIKLKGANFYARVVKDDIDCLACGLDNENAEVRKARLKIPIVRGYIGE

CVKKNKMLPDLYLKLENALESKSTVTYVKVKSVAHVESGLQDTAHILEDOKSYN

ATLNMSDLALGVNSYVYLQITIEDQDGGSECVYFRKRWGVGSEKIGQKLEMSKTEAIK

EFKRLFEKTKGNSEWAECKTNRKQRFYPLDVGKVKAPRKDIEMSKSLAPQ

LLELMKMLFNVTYIRAAWEPEINNSEMPLEKLSKENIEKFEALTIEIQLLKDTAQ

ALAVRESLIVAASNRFTLIPS.IHPHIIIRDDEDLWKAKMLEALQDIEIAKIVGFS

DSDESDDKYMKLHCDITPLAHDSEDKLIEQYLLNTHAPTHDKMSLELEVESLDR

GELNKSRYKNNHNNKMLLWHLGSLRNFVGLLSOGLRLAPPEAPVDTGYMGKGLYFAD

LVSKSAQCYVDNRNNPVGKMLLSEVALGMDKATSMKPPKRGKSTKGLKTVPL

ESEFVKWRDDVVVPCGKVPVSSIRSELNNEYIVYNTSQVKMQLLKVPHHRR"

BASE COUNT 968 a 604 c 813 g 826 t
 ORIGIN
 alignment_scores:
 Quality: 4666.50 Length: 985
 Ratio: 4.980 Gaps: 6
 Percent Similarity: 95.127 Percent Identity: 94.416
 alignment_block:
 US-09-236-995D-2 x ZMPARP2
 Align seg 1/1 to: ZMPARP2 from: 1 to: 3211
 1 MetAlaAlaProProLysAlaTrpLysAlaGluTyrAlaLysSerGlyAr 17
 113 ATGGCGGGCGCCCAAGGCGTGGAGGGGGAGTAGTCCCAAGTCTGGGGC 162
 17 gAlaSerCysLysSerCysArgSerProLleAlaLysAspGlnLeuArgL 34
 163 GGCGCTCGTCAAGTCATGCGGTCCTTATCGCCAAAGGACCAGCTCCGTC 212
 34 euGlyLysMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrp 50
 213 TTGGCAAGATGGTTCAGGCGTCACAGTTCGACGGCTTCATGCGGATGG 262
 51 AsnHisAlaArgCysIlePheSerLysLysAsnGlnIleLysSerValAs 67
 263 AACCATGCC.....AGCGTTGA 279
 67 pAspValGluGlyIleAspAlaLeuArgTrpAspGlnGluLysIleA 84
 280 CGATGTTGAAGGGATAGATGCACCTTAGATGGGATGATCAAGAGAAGATAC 329
 84 rGAsnTyrValGlySerAlaSerAlaGlyThrSerThrAlaAlaPro 100
 330 GAACACTACCTGGGAGTGCTCAGCTGGTACAAAGTTCTACAGCTGCTCCT 379
 101 ProGluLysCysThrIleGluIleAlaProSerAlaArgThrSerCysAr 117
 380 CCTGAGAAATGTACAATTGAGATTGCTCCATCTGCCCGTACTTTCATGTAG 429
 117 qArgCysSerGluLysIleThrLysGlySerValArgLeuSerAlaLysL 134
 430 ACGATGACGTAAGAAAGATTACAAAAAGGATCGGTCGCTCTTCAGCTAAGC 479
 134 euGluSerGluGlyProLysGlyIleProTrpTyrHisAlaAsnCysPhe 150
 480 TTGAGATGAAGGTCCCAAGGGTATACCATGGTATCATGCCAAGTCTTC 529
 151 PheGluValSerProSerAlaThrValGluLysPheSerGlyTyrAspTh 167
 530 TTTGAGGTATCCCGCTCTGCAACTGTTGAGAAGTTCTCAGGCTGGGATAC 579
 167 rLeuSerAspGluAspLysArgThrMetLeuAspLeuValLysLysAspV 184
 580 TTTGTCCGATGAGGATAGAAACCATGCTCGATCTTGTAAAAAAGATG 629
 184 aGlyAsnAsnGluGlnAsnLysGlySerLysArgLysLysSerGluAsn 200
 630 TTGGCAACAATGACAAAAATAAGGTTTCCAGCGCAAGAAAAAGTGAAT 679
 201 AspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrSerGluG 217
 680 GATATTGATAGCTACAAATCCGCCAGGTAGATGAAAGTACATCTGAAG 729
 217 yThrValArgAsnLysGlyGlnLeuValAspProArgGlySerAsnThrS 234
 730 TACAGTGGCAACAAGGCAACTTTGTAGACCCACGCTGGTTCCCAATACTA 779
 234 erSerAlaAspIleGlnLeuLysLysGluGlnSerAspThrLeuTrp 250
 780 GTTCAGCTGATATCCAACTAAAGCTTAAGGAGCAAAAGTACACACTTTGG 829

251 LysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGluLeuArgAs 267
 830 AAGTTAAAGGATGGACTTAAGACTCATGTATCGGCTGCTGAATTAAGGA 879
 267 pMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArgHisLeuL 284
 880 TATGCTTGAGGCTAATGGCAGGATACATCAGGACCAGAAAGGACCTAT 929
 284 euAspArgCysAlaAspGlyMetLeuPheGlyAlaLeuGlyProCysPro 300
 930 TCGATCGCTGCGGATGGAATGATATTGAGCGCTGGGCTCTTGCCCA 979
 301 ValCysAlaAsnGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerG 317
 980 GTCTGTGCTAATGGCATGTACTATTATTAATGCTCAGTACCAATGCAGTG 1029
 317 yAsnValSerGluTrpSerLysCysThrTyrSerAlaThrGluProValA 334
 1030 TAATGTGTACAGTGGTCCAAGTGTACATCTCTGCCACAGAACCTGTCC 1079
 334 rGValLysLysLysTrpGlnIleProHisGlyThrLysAsnAspTyrLeu 350
 1080 CGGTTAAGAGAAGTGGCAAAATTCACATGCAACAAAGAATGATTACCTT 1129
 351 MetLysTrpPheLysSerGlnLysValLysLysProGluArgValLeuPr 367
 1130 ATGAAGTGGTTCAAAATCTCAAAAGGTTAAGAAACCCAGAGAGGTTCTTCC 1179
 367 oProMetSerProGluLysSerGlySerLysAlaThrGlnArgThrSerL 384
 1180 ACCAATGTCACTCAGAAATCTGGAAGTAAAGCAACTCAGAGAACATCAT 1229
 384 euLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValGlyGln 400
 1230 TGCTGTCTTCTAAAGGTTGGATAAATTAAGGTTTCTGTTGAGGACAA 1279
 401 SerLysGluAlaAlaAsnGluTrpIleGluLysLeuLysLeuAlaGlyAl 417
 1280 TCAAAAGACAGCAAAATGAGTGGATTGAGAAGCTCAAACTTCTGCTGTC 1329
 417 aAsnPheTyrAlaArgValValLysAspIleAspCysLeuIleAlaCysG 434
 1330 CAACCTTCTATGCCAGGTTGTCAAAGATATTGATTGTTTAAATGCATGTG 1379
 434 LyGluLeuAspAsnGluAsnAlaGluValArgLysAlaArgArgLeuLys 450
 1380 GTGAGCTCGACAAATGAAATGCTGAGTCAAGAAAGCAAGAGGCTGAAG 1429
 451 IleProIleValArgGluGlyTyrIleGlyGluCysValLysArgThrL 467
 1430 ATACCAATTTAAGGGAGGTTACATTTGGAGAATGTTTAAAGAAACAA 1479
 467 yScCysCysHisLeuIleCysIleAsn..TrpAsnAlaLeuGluSerSerL 483
 1480 AATGCTGCCATTTGATTGTTATTAACCTAGAGAATGCTTTAGAGTCTCAA 1529
 483 ySGly***ThrValThrValLysValLysGlyArgSerAlaCysSer*** 499
 1530 AAGCGAGTACTGTCATCTGTTAAAGTTAAGGCGCAAGTGTCTTTCATGAG 1579
 500 SerPro***ValCysLysAsnThrAlaHisIlePro****TrpGluLys 516
 1580 TCCTCTGGTGTG...CAAGATACCTGCTCACATTTCTTGA..AGATGGGAAA 1625
 516 sHisIleGlnCys***LeuLysHisValLeuThr***His***ValCys.. 532
 1626 GCATATACAAATGCACCTTAAACATGCTCGACCTGGC...ACTAGGTGTG 1672
 533 ThrGlyTyrTyrValLeuGlnIleIleGluGlnAspGlySerGluCys 549
 1673 AACAGCTACTATGCTACCTCAGATCATTTGACAGGATGATGGGCTGAGTG 1722

549 sTyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGly 566
1723 CTACGTATTTCTAGTGGGACGGTGGGAGTCAGAAAAATTGAGGGC 1772
566 InLysLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysArg 582
1773 AAAAAGCTGGAGGAGATGTCAGAAATGAGGCAATCAAGGAATCAAAAGA 1822
583 LeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysTh 599
1823 TTATTTCTTGAGAGACTGGAACATCATGGGAAGCTGGGAATGTAAAC 1872
599 rAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrG 616
1873 CAATTTTCGGAAGCAGCCTGGGAGATTTTACCCACTTCATGTTGATTG 1922
616 yValLysLysAlaProLysArgLysAspIleSerGluMetLysSerSer 632
1923 GTGTTTAAAGAAAGCACCAAAACGGAAGATATCAGTGAATGAAAAGTTCT 1972
633 LeuAlaProGlnLeuLeuGluMetLysMetLeuPheAsnValGluTh 649
1973 CTTCCTCTCAATTCCTAGAACTCATGAAGATGCTTTTCAATGTGGAGAC 2022
649 rTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetProL 666
2023 ATATAGAGCTGCTATGATGGAATTTGAAATTAATATATGCAGAAATGCCTC 2072
666 euGlyLysLeuSer*****AsnIleGlu***GlyPheGluAlaLeuThr 682
2073 TTGGGAAGCTAAGCAAGGAAATATGTGAAGAGGATTTGAAGCATTAAC 2122
683 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu** 699
2123 GAGTACAGCAATTTATT.GAGGACACCGCTGATCAAGCACTGGCTG.TT 2170
699 *gluLysAla***LeuLeu*****PheSerLeuLeuSerL 716
2171 AGAGAAAGCTTAATTTGTTGCTGCGAGCAATCGCTTTTTCACCTTATCCC 2220
716 euLeuPheIleLeuIleLeuTyrGlyMetArgMetIleSerTyrSerLys 732
2221 TTCTATTATCTCATATTATACGGGATGAGGATGATTG.ATGATCAA 2269
733 AlaLysMetLeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleVa 749
2270 CGGAAATGCTTGAAGCTCTCGAGATATTGAAATTCCTCAAGATAGT 2319
749 lcyPheAspSerAspSerAspGluSerLeuAspLysTyrMetLysL 766
2320 TGGCTTCGATAGCGACAGTGAATCTTTGATGATATAATATATGAAC 2369
766 euHisCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeu 782
2370 TTCACGTGACATCAACCCCGCTGGCTACGATAGTGAAGATTACAAGTTA 2419
783 ileGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSe 799
2420 ATTGAGCAGTATCTCTCAACACACATGCTCTACTCACAGGACTGGTC 2469
799 rLeuLeuLeuGluValPheSerLeuAspArgAspGlyGluLeuAsnL 816
2470 GCTGGAAGCTGGAGGAAGTTTTCCTACTGTATGATCGAGATGGAGAACTTA 2519
816 yTyrSerArgTyrLysAsnLeuHisAsnLysMetLeuLeuTrpHis 832
2520 AGTACTCAAGATATAAAATATCTGCAATAACAAGATGCTATTATGGCAC 2569
833 GlySerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIl 849
2570 GGTTCAGGTTGACGAAATTTGTGGGAATTTCTTAGTCAAGGCTAAGNAT 2619
849 eAlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeu 866

2620 TGACACCTCTGAGGCACCTGTTACTGGCTATATGTTGGCAAGGCTCT 2669
866 yPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArg 882
2670 ACTTTGGAGATCTAGTAAGCAAGAGCGCACATACTGTTATGTGGATAG 2719
883 AsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMe 899
2720 AATAATCTCTAGGTTTGATGCTCTCTTCTGAGGTTGCTTTAGGAGACAT 2769
899 tTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysH 916
2770 GTATGAACCTAAAGAAACCCACGTCATGGACAAACCTCCAAGAGGGAAC 2819
916 isSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 932
2820 ATTTCAGCAAGGATAGGCAAAACCGTCCACTGGAGTCAGAGTTTGTG 2869
933 LysTrpArgAspValValValProCysGlyLysProValProSerSe 949
2870 AAGTGGAGGATGATGCTGCTGCTGCGCAAGCGGTGCATCATC 2919
949 rIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrS 966
2920 AATTAGAGCTCTCAACTCATGTACATGAGTACATCGTCTACACACAT 2969
966 erGlnValLysMetGlnPheLeuLysValArgPheHisLysLysArg 982
2970 CCAGGTGAAGATGACGATCTTCTGCTGAAGTGGCTTCCATCAAGAGG 3019

seq_name: gb_pl:ATH131705

seq_documentation_block:

LOCUS ATH131705 3187 bp mRNA PLN 18-DEC-1998
DEFINITION Arabidopsis thaliana mRNA for poly(ADP-ribose) polymerase.

ACCESSION AJ131705

VERSION AJ131705.1 GI:4038490

KEYWORDS NAD(+) ADP-ribosyltransferase; parp-1 gene; poly(ADP-ribose) polymerase.

SOURCE thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 3187)

AUTHORS Kazmaier,M.

TITLE Direct Submission

JOURNAL

Submitted (17-DEC-1998) Kazmaier M., Departement D'Ecophysiologie Vegetale Et de Microbiologie, Commissariat A L'energie Atomique, CEA Cadarache, Lab. De Radiobiol. Vegetale, Bat. 185, 13108 St. Paul-les-Durance, FRANCE

REFERENCE 2 (bases 1 to 3187)

AUTHORS Doucet-chabeaud,G. and Kazmaier,M.

JOURNAL

Unpublished

FEATURES

source

Location/Qualifiers

1..3187

/organism="Arabidopsis thaliana"

/cultivar="landsberg erecta"

/db_xref="taxon:3702"

11..2963

/gene="parp-1"

11..2963

/gene="parp-1"

/standard_name="NAD(+) ADP-ribosyltransferase"

/EC_number="2.4.2.30"

/function="ADP-Ribose polymer synthesis"

/note="mol. weight: 111 kDa"

/codon_start=1

/evidence=experimental

/product="poly(ADP-ribose) polymerase"

/protein_id="CAA10482.1"

/db_xref="GI:4038491"

1313 ACAAGCTGTTTGGTTGTTGGCTGACAGATATCAGAGACGCTGAAT 1362

443 lArgLysAlaArgLeuLysIleProIleValArgGluGlyTyrIleG 460

1363 GAGAAAGCAAGGAGGATGAAAGTGGCAATCGTGAGAGAGGATTATTTGG 1412

460 lYGluCysValLysArgThrLysCysCysHisLeuIleCysIleAsnTrp 476

1413 TTGATTCGTTTAAAAACACAGAGG.....AAAGTCCATTTTGACAAG 1453

477 AsnAlaLeuGluSerSerLysGly****ThrValThrValLysValLysGI 493

1454 TACAAAATTTAGACACACTAGTGAGAGCCCTGTCACTGTTAAAGTAAAGG 1503

493 yArgSerAlaCysSer***SerPro***ValCysLysLysAsnThrAlaHisI 510

1504 ACGAAGGCGT.....GTGCATG 1520

510 lePro*****TrpGluLysHisIleGlnCys**Leu..... 522

1521 AAGCGCTGGCCTCCAAGAGCAC.....TGTCACATCTTGAAGATGGG 1564

523LysHisValLeuThr***His***ValCysThrGly.. 534

1565 AACGATATCTATAACACAACACTCGAGCATGCTGTACTCTCTACCGGTAT 1614

535TyrThrValLeuGlnIleIleGluGlnAspAspGlySerGluC 549

1615 CAATAGTTATTACATATCTCCAGATAATCAAGAGAGATAAAGTTTCAGATT 1664

549 ysThrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGly 565

1665 GTTACGTATTTCGTAATGGGCGCGAGTTGGAAATGAAAGATGCTGTGGT 1714

566 GlnLysLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysTr 582

1715 AACAAAGTGGAGGAATGTCAAAAGTCTGATCGCGGTTCCAGAAATCAACG 1764

582 gLeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLys 599

1765 TCTATTCTTGAAAAACCGGAACACATGGGAATCTTGGGACAAAAA 1814

599 hrAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyr 615

1815 CGAATTTCCAGAAACACCTCGAAATTTCCCGGTTGGACATGTGATTAT 1864

616 GlyValLysLysAlaProLysArgLysAspIleSerGluMetLysSerSe 632

1865 GGAGTTAATAACCAAGTAGCCAAAAAAGAGCCATTTCAGACCAGTAGCAA 1914

632 rLeuAlaProGlnLeuLeuGluLeuMetLysMetLeuPheAsnValGluTr 649

1915 CTTGCTCCATCATTAATAGAAATTGATGAACATGCTTTTGTATGTGAAA 1964

649 hrTyrArgAlaAlaMetMetGluPheGlu****AsnMetSerGluMetPro 665

1965 CTTACAGATCTCGAATGATGGAGTTTCGAGATAAATATGTCAGAGATGCCA 2014

666 LeuGlyLysLeuSer*****AsnIleGlu****GlyPheGluAlaLeuTh 682

2015 CTTGGGAAGCTCAGCAAAACATAATATACAGAAGGGTTTTTGAGGCATTTC 2064

682 r*****LeuPheGluGlyHis.Arg***SerSerThrGlyLeu 698

2065 GGAGATACAGAGCTATTG.....ACTGAAGCGACCCCGAGCTACTA 2108

699 **GlyLysAla***LeuLeu*****PheSerLeuLeuSe 715

2109 TGAAGAAAGACTGCTGTGTGATGCTAGTAAACAGATTTTTTACCATGATC 2158

715 rLeuLeuPheIleLeuLeuTyrGlyMetArgMetIleSerTyrSerL 732

2159 CCTTCTATTATTCCTCATATATTCGAGATGAAGTAGACTTT.AAGTCAA 2207

732 ysAlaLysMetLeuGluAlaLeuGlnAsnPilecIuIleAlaSerLysIle 748
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2208 AGGTCAAATGCTCGAGGCCTTCAGGATATCGAAATAGCTTCAAGAATA 2257

749 ValGlyPheAspSerAspSerAspGluSerLeuAspAspLysTyrMetLy 765
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2258 GTTGCCCTTGATGTGATACACCGGAATCCTTGATGATTAAGATTAAGA 2307

765 sLeuHiCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysL 782
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2308 GCTGATTGCCGATATCTCACACTTCTCATGATACCGAAGATTATCGAT 2357

782 euIleGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrp 798
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2358 TAATCGAAGATATCTTTAACACAACATCATGCCCAACGCATACAGATGG 2407

799 SerLeuGluLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAs 815
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2408 AGTCITTAGCTAGAGGAAGTTTTTGCTCTTGAAGAAGAAAGGAGGTTCG 2457

815 nLysTyrSerArgTyrLysAsnLeuHisAsnLysMetLeuLeuTrpH 832
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2458 TAAATATGCTCCCACACAGAGAAAACCTGGCANTAGACTCTCCTATGCG 2507

832 tsGlySerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArg 848
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2508 ATGGTTCTCGAATTAAACGAATTTTGTGGAATATTTGAACCAGGACTGAGA 2557

849 IleAlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLe 865
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2558 ATTGCACCTCCAGAAGCTCTGCTACTGGTTTACATGTTTGGAAACGGAT 2607

865 utYrPheAlaAspLeuValSerLysSerAlaGlnTyrCystYrValAspA 882
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2608 AVACTTTGCTGACCTTGTTCAGTAAAGTGCTCAGTACTGCTACACTTGTA 2657

882 rgAsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAsp 898
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2658 AGAAAAATCCGGTGGGTCTAATGCTTCTGAGTGAAGTGCATTTGGGAGAA 2707

899 MetTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLy 915
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2708 ATACATGAGCTAACAAAAAGCTAAGTATATGATGAANAACCTCCGAGAGGAA 2757

915 shisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheV 932
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2758 ACATCGACCAAGGCTCGGCAAGAAAGTGCCTCAAGATTCCGAGTTTG 2807

932 alLysTrpArgAspValValProCysGlyLysProValProSer 948
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2808 CCAAGTGGAGAGGTGATGTGACTGTCTCCTGTGGAAACCTGTTTCATCA 2857

949 SerIleArgSerSerCluLeuMetTyrAsnGluTyrIleValTyrAsnTh 965
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2858 AAGGTCGAAGGCTCTGAGCTTATGTACAAATGAGTATATCTGCTACGATAC 2907

965 rSerGlnValLysMetGlnPheLeuLeuLysValArgPheHisLysA 982
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2908 AGCCAGGTTGAAGTTCGAGTTCTTCTGTTGAAAGTAAGGTTTAAAGCACAGA 2957

982 rg 982
||
2958 GA 2959

seq name: qb_pl:AC006593

| | | | |
|--------------------------|--|------------|-------------|
| seq_documentation_block: | | PLN | 05-APR-2000 |
| LOCUS | AC006593 | 79663 bp | DNA |
| DEFINITION | Arabidopsis thaliana chromosome II section 176 of 255 of the complete sequence. Sequence from clones F16D14, T28P16. | | |
| ACCESSION | AC006593 | AE002093 | |
| VERSION | AC006593.4 | GI:6598616 | |

KEYWORDS
SOURCE
ORGANISM

HTG.
thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 79663)

REFERENCE
AUTHORS

Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D.,
Fuji, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E.,
Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Unayam, L.,
Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H.,
Gordon, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
Venter, J.C.

TITLE

Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana

JOURNAL
MEDLINE
PUBMED
REFERENCE

Nature 402 (6763), 761-768 (1999)

20083487

10617197

2 (bases 1 to 79663)

Lin, X.

Direct Submission

JOURNAL
COMMENT

Submitted (09-Mar-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:4432811.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(<http://www.tigr.org/tdbat/at.html>).

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(<ftp://arthur.epm.ornl.gov/pub/xgrail>), GeneFinder (Phil Green,
University of Washington), Genscan (Chris Burge,
<http://genomic.stanford.edu/genscanw.html>), and NetPlantGene
(<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curations based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were
identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
and Satoshi Tabata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Lixin
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES
source

Location/Qualifiers
1. .79663

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

1. .79462

/note="Sequence from clone F16D14"

misc_feature

mRNA

complement(join(<1349..1687,1764..1967,2138..2263,
2342..2623,2709..2879,2963..3124,3898..3961,4063..>4406))
/gene="At2g31170"

gene

complement(<1349..>4406)

CDS

/gene="At2g31170"

/note="F16D14.1"

complement(join(1349..1687,1764..1967,2138..2263,
2342..2623,2709..2879,2963..3124,3898..3961,4063..4406))

/gene="At2g31170"

/codon_start=1

/product="putative cysteinyl-tRNA synthetase"

/protein_id="AAD20662.1"

/db_xref="GI:4432812"

/translation="MASSVINLFKSCRPETPIRFSLSPKSQRIQFIPLRPGKETOLRR
CFTTSSLDGGAPISGGKELHLNHTMSKKELFKPKVEGKVMYVCGVTAYDLSHIG
HARVYVFDVLLRYLHLGYEVSYYRNFVDVDDKIARAKELEEDPISLRRFCSEFN
RMEQLQCLDPSQVPRVSDHIPOIIDLIOIIONGVYKVDGDIYFVDKPTCYGLS
GRKLEDNRAGERVAVDTRKKHPADFAIMKAKEGPFWESPMGRGPGWHIECSAMSA
AYLGYSFDIHGGMDLVFPHEHEIAQSCAACDSSNIYWHNGFVTVDSEKMSKLG
NEFTTROVDLHPLALRFLMGTHIRSPINYSDFLLESASERIFYIOTLHDCESAL
GKSDSTFENGVSDDTLTSINTFRTEFVMSDDLLTPVTAAAMSEPKTINDLHTTR
VGKQARREESKALETTIRVDLITILGMPTSYSEVLQEKALKRAGLEEDVLQR
VOERTDARKKEYERSDAIKDLAKVGIALMSDEGTTWRPAIPLALQEPVTTTP"
complement(join(5525..6082,6603..6732,6833..>6965))
/gene="At2g31180"

gene

complement(5525..>6965)

CDS

/gene="At2g31180"

/note="F16D14.2"

complement(join(5596..6082,6603..6732,6833..6965))

/gene="At2g31180"

/codon_start=1

/product="putative MYB family transcription factor"

/protein_id="AAD20663.1"

/db_xref="GI:4432813"

/translation="MGRAPCEKMGVGRGPTPBDQILINVIHYLGHNSWRALPKHA
GILLRGCKSLRWLNILRPDIKRGFTPOEEQTTINLHESLGNWSATAAKLPQRTDN
ETKNVHTHLKRLSKNLNGDITKDVNGINETTNEKGSIVDTASIQOEPSNITTF
DISDNKDDIMSYEDISALIDDSFSDSVISYDNSKNKKEKKEGLIDRNSKKCSYS
NSKLYNDDEWFDFVFTSNRIEFSIDPEF"
complement(7897..7920)

repeat_region

/rpt_family="POLY_A"

8038..8151

/rpt_family="TAAAAn"

9946..9977

repeat_region

/rpt_family="TAAAn"

complement(join(<10248..10673,10764..10886,10983..11162,
11326..11423,11506..11599,11690..11749,11871..11955,
12057..>12145))

gene

/gene="At2g31190"

complement(<10248..>12145)

/gene="At2g31190"

/note="F16D14.3"

complement(join(10248..10673,10764..10886,10983..11162,
11326..11423,11506..11599,11690..11749,11871..11955,
12057..12145))

CDS

/gene="At2g31190"

/note="unknown protein"

/codon_start=1

/protein_id="AAD20664.1"

/db_xref="GI:4432814"

/translation="MKVVDARVPQKRVESFLNKKFPSPGYXSVNBEGLRYTQFRAL
QHSAAALSVLSTQSLFLAAGLRPTPAQATVWSWILKDGQHVKGKLTCSLGMADSE
PKRWILADVLDTGLGVLPCLPFLHLEAGLGNFAGMATVAARATRLPIYSF
AKEGNLSDFAKGAIATLNFVAGIGAGIOLASTICSSMEKGLVGSILSVHVSVV
EOMEGVPTINLPORALIVANFLKTKGVSFDPDLPOEDLMPEPIQDAGNVKVR
ALHKAVKPSVORLKVVEEKLHSGKSWTDLVLEHDATGADALRGWLVAAYVKSM
TKIYNDDDDIILQDAYDKMNDVNFPLSQVQAKGYTDFLDGTGTFFAM"
join(12936..12984,13233..13498,13588..13809)

mRNA

/gene="At2g31200"

12936..13809

gene

/gene="At2g31200"

/note="F16D14.4"

join(13251..13498,13588..13738)

CDS

/gene="At2g31200"
/codon_start=1
/product="actin depolymerizing factor 6"
/protein_id="AAD20665.1"
/db_xref="GI:4432815"
/translation="MGVADESKTTFLELQKKTHRVVFKIDESKEVVVEKTNPT
SYDDFLASLPDNCRYAVDFDFTVSENCOKSKIFFFAWSPSGIRAKVLYSTSKDQ
LSRELOGIHVEIQDPTDTEVDLVRERAN"
join(<15445..16170,16266..>16739)
/gene="At2g31210"
/note="F16014.5; predicted by genefinder"
join(15445..16170,16266..16739)
/gene="At2g31210"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD20666.1"
/db_xref="GI:4432816"
/translation="MYEESCDFDPSNVNNGGFCFAEITFTVSHQFOPPLGSGTNSF
DDDLKPTMDFSVFVPSVILPNSSETQNISSNNHILNOMIQESNMGVSEDNSNPFM
NTSHPTTTPDIPOLLSLHLPRCSMSLPSDIMGSCFTYDPLFLNLPPQPLPS
NDYSYLIDTNTTQKDESNGDNNNAQFDSGIIIEFSKEIRKRGKRGKRPFT
ERERCHLNERVEALKLLIPSPSGDRASILQDGDVINEELRRVSELKYLVERKRG
GRKNEVDNNKLLDHDGNEDDDDDDMMKKPESDVIDOCSSNNSLRCSWLQK
SKVTEVDRIIVDDEVTIKVOKKINCLLLVSKVLDQLDLHLHVAGQIGEHVSFLF
NTKVSQD"
complement(16419..16456)
/rpt_family="CAT)n"
20630..20718
/rpt_family="TAAA)n"
join(21922..22723,22818..23267,23350..>23451)
/gene="At2g31220"
21922..>23451
/gene="At2g31220"
/note="F16014.6"
join(21932..22723,22818..23267,23350..23451)
/gene="At2g31220"
/note="unknown protein"
/codon_start=1
/protein_id="AAD20667.1"
/db_xref="GI:4432817"
/translation="MGCFDPTPAEVTVESSFSQAEPPLPPQVLVAGSTNSNCVSE
VLESEFLSPDQCAQSTPLQPHINPPPPPCDOLHNNLIHOMASHOQOHSNWD
NGTDFVNLGPNATPDLISLHLPRCSLPENHPSSMLPTSPSDIMSSSSAAAVMY
DPLFLHLPMPQDQNLKRGSCLLGVEDQIQMDANGMVLVPEGANNGGFENEI
LEFNGVTRKGRSKRSPTSPERRRVHFNDRFDLANLIPNPKIDRASIVGEAID
YIKELLRTIEEFKMLVERKCRFRKRARVGGGGEDQEEEDTVNRYPOSEVDQ
SCFNKNNNSLRCSWLKRSKVTEVDVRIIDDEVTIKLVQKKINCLLFTTKVLQDLO

mRNA
gene
CDS
repeat_region
repeat_region
mRNA
gene
CDS
alignment_scores
Quality: 1973.50 Length: 1786
Ratio: 2.501 Gaps: 42
Percent Similarity: 44.177 Percent Identity: 30.571
alignment_block
US-09-236-995D-2 x AC006593/rev ..
Align seg 1/1 to reverse of: AC006593 from: 1 to: 79663
1 MetAlaAlaProLyAlaTriPhyAlaGluTyAlaLysSerGlyAr 17
78368 ATGGCAAGCCACATAGCCGTGGAGCGGAGGATATGCAAAAGTCGCGAG 78319
17 gAlaSerCysLysSerCysArgSerProLleAlaLysAspGlnLeuArgL 34
78318 GTCTTCATGTAACATTCAGTCGTCGTCATTAACAAGGAGAACTTTCGTC 78269
34 euClyLysMetValGlnAlaSerGlnPheAspGlyPheMetPro..Met. 49
78268 TTGGAAAGTTGGTTCAATCTACTACTCTCGATGGCATCATGCCCGATGA 78219
49

77318 ACTTCGTTGGGAAGATCAGCAAAAGATTAGAAAATATGTCGAATCTGGAG 77269

91 e rAlaGlyThrSerSerThrAlaAlaProProGluLysCysThr..... 105

77268 CAGGAGTTAACACACAGCACAGACAGGCACAGCAGCAGCAGTACCGCT 77219

106l leGluLeAlaProSerAlaArgTh 114

77218 ANTAAATGCCAAACTAGAAATATGGATTTGAAGTTGCACAACTTCTCGNGC 77169

114 rSerCysArgArgCysSerGluLysTleThrLysGly..... 126

77168 CGGTTGCAGAAAGCTGTAGCGAAAGATCTTGAAGAGAGAGTAATGTT 77119

126 126

77118 GTTCTACTTCATCTGTCATGTTGTAGCGTTTATATGAACCTGTTCCGAT 77069

127Ser.ValArgLeuSerA 132

77068 ATCCATTAGGCTTTGTTTGCATTTTGTTCAGGTACGTATATTC 77019

132 laLysLeuGluSerGluGlyProLysGlyIleProTptTyrHisAlaAsn 148

77018 CCAAGCCTGAAGCCCGGTAAACAAAGCTTTGATCTGCACTACGCCTAAA 76969

149 CysPhePheGluValSerProSerAlaThrValGluLysPheSerGlyTr 165

76968 TGTTCCTTGAATGTCTCTCTACTCACTGAACCTGGAAAGTTGTCTGGATG 76919

165 pAspThrLeuSerAspGluAspLysArgThrMetLeuAspLeuValLysL 182

76918 GAGAATATACCAGACTCAGACCAAGAAGCTCTCTCCCTTAGTGAAGA 76869

182 ysAspValGlyAsnAsnGluGlnAsnLysGlySerLys..... 194

76868 AAGCTCTCGCGCAGCCAAAACCTCGTAAGAGTCTTAAGGTGTACGGTAAT 76819

195ArgLysLysSerGluAsnAspIleAspSerTyr..... 205

76818 ATATTGATTTCTCTTTACAGGATCCCTGATTTGCAGTACTTTAGTCTGAT 76769

206LysSerAlaArgLeuA 211

76768 CTTTCCGTTAATTTATTTTGGCCCACTGGGACAGACAGCAGCAAGAC 76719

211 spGluSerThrSerGluGlyThrValArgAsnLysGlyGlnLeuValAsp 227

76718 GTCAAACAAATTAAGAGCAGGCACAAACGAAATAATGATTCTGTTGAT 76669

228 ProArgGlySerAsnThrSerSerAlaAspIleGlnLeu..... 240

76668 AACGAGAAGTCGAAACTAGCAAAAGTAGTTTTTGACATGCTTACAAGTGG 76619

241LysLeuLysGluGlnSerAspTr 248

76618 TGCCTTACAACCTTGTAGCAAGAAAGAAATGGAGGCCCAAACTAAGG 76569

248 hrLeuTrpLysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGlu 264

76568 AATTGTGGGACCTCAGAGGATGATCTGAAATAATATGTAACATCAGCTGAG 76519

265 LeuArgAspMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluAr 281

76518 TTGCGGAAATGCTTGAAGTAAATGAACAAAGATACAAGAGGATCTGAACT 76469

281 gHisLeuLeuAspArg..... 286

76468 TGATCTCGGTGATAAATGTTAAATTTACTTTGCGAATCTGTCTGGAAGTTA 76419

286 286

76418 ATTGCTTGCTTTTGTTCATTATTTCTGTACTGTTGTGGATCTTCTTAA 76369

286 286

76368 AGGTAATTGGATTAGAGAAACCGATGCATAATCATATGCCTAACCTCTC 76319

287Cys..AlaAspGlyMetLeuPheGlyAlaLeuGlyPr 298

76318 TATTGTTTCTATTAGTCTGAAGCATGATGTTTGGCCCACTGCCTCT 76269

298 oCysProValCysAlaAsnGlyMetTyrTyrAsnGlyGlnTyrGlnC 315

76268 CTGCCCAATGCTCTGGGCATCTTCTTCCTCCGAGGACTTTACCGAT 76219

315 ySerGlyAsnValSerGluTyrSerLysCysThrTyrSerAlaThrGlu 331

76218 GCCATGTGACATCTTCAGAATGGACAAATGTCATCTCCACTTTGGAT 76169

332 ProValArgValLysLysTyrGlnIleProHisGlyThrLysAsnAs 348

76168 CCAGCCGCATCAAGGAAGTGGAAATCCCTGACGAAACAGAAAAATCA 76119

348 pTyrLeuMetLys..... 352

76118 ATTCTCTTCAAGGTAAACCACTCTTTTGTAGAGTTGATATTGGTATGG 76069

352 352

76068 GTTAGGTTAGATGGGTATATATTACTCATCAAGTTAATATTATATTTT 76019

353TrpPheLysSerGlnLysValLysLysProGluArgVa 365

76018 GATTACTTCGACGTGAATAAGTTCAAAAGAGTGTGAAGCCAAACGAT 75969

365 lLeu...ProPrometSerProGluLysSer...GlySerLysAlaThrC 380

75968 TCTGGCTCCTGATTGTCTGGCAGACATCTCAGGGTCAAGGTTCTAAAG 75919

380 lArgThrSerLeuLeuSerSerLysGlyLeuAspLysLeuArgPheSer 396

75918 ATGCAACTGACTCTCTCAAGGAGTCAAAGGCTAGCAGATCTTAAAGTTTCA 75869

397 ValValGlyGlnSerLysGluAlaAla..... 405

75868 ATTGCTGGAATATCTAAGAAAGGATATGTTCTCTGTCAAAAATATAT 75819

406AsnG 407

75818 TGTAGACTATCTTTGCACCTTCATTGCTTGATGGATCATTTTCCACCAAC 75769

407 lUTriPleGluLysLeuLysLeuAlaGlyAlaAsnPheTyrAlaArgVal 423

75768 CATGGAAGAAGAGAATTCAGAGAACTGGTCAGAGTTTCATGCTAATGTT 75719

424 ValLysAspIleAsp..... 428

75718 AAAAAGGTATGGATACTCGAGTGTCATATTTTGATATTATGTGCACACTA 75669

429Cys.....LeuI 431

75668 GTTTTAGCTTTTGAATATCTCTCTTCACAAAATTTGATGGTACAAGCTGT 75619

431 leAlaCysGlyGluLeuAsp.....AsnGluAsnAla 441

75618 TTGGTTGTTTGGACTGACAGATATCAGAGACGCTGGAATGAGAAAGGC. 75569

442 GluValArgLys..... 445

75568 AAGGTAGAATATTGCCTTCATTGTGCTGATACTCTTTAACTTTGTCACAC 75519

446AlaArgArgL 449

75518 TATTTTTCACATTGAACGCTGGAGTTGATAAATGTGTGGCAGAGGA 75469

```

449 euYsilleProIleValargGluGlyTyrIleGlyGluCysVallysArg 455
    :::: ||||| ||||| ||||| :::: ||||| |||||
75468 TGAAGTGGCAATCGTCAGAGAGATTAATTTGGTTGATGTTTAAAAAA 75419
    :::: ||||| ||||| ||||| :::: ||||| |||||
466 ThrLysCysCysHisLeuIleCysIleAsnTrpAsnAlaLeuGluSerSe 482
    :::: ||||| ||||| ||||| :::: ||||| |||||
75418 CAGAGG.....AACTTCCATTTTGACAAGTACAAAATTCAGACAC 75378
    :::: ||||| ||||| ||||| :::: ||||| |||||
482 rLYSGly****ThrValThrVallysVallysGlyArgSerAlaCysSer* 499
    :::: ||||| ||||| ||||| :::: ||||| |||||
75377 TAGTCAGAGCCTTGTCTACTGTAAAGTAAAGGACGAAGCGCT..... 75335
    :::: ||||| ||||| ||||| :::: ||||| |||||
499 **SerPro**valCysLysAsnThrAlaHisIlePro*****TrpGlu 515
    :::: ||||| ||||| ||||| :::: ||||| |||||
75334 .....GTGCATGAAGCGTCTGGCGCTCCAA 75311
    :::: ||||| ||||| ||||| :::: ||||| |||||
516 LysHisIleGlnCys**Leu.....LysHi 524
    :::: ||||| ||||| ||||| :::: ||||| |||||
75310 GAGCAC.....TGTCACTCTTTGACAGTGGGAACAGTATCTATAACAC 75267
    :::: ||||| ||||| ||||| :::: ||||| |||||
524 sValLeuThr***His***ValCysThrGly.....TyrTyrValL 538
    :::: ||||| ||||| ||||| :::: ||||| |||||
75266 AACTCTGAGCATGTCGTATCTCTACCGGTATCAATAGTTATACATAC 75217
    :::: ||||| ||||| ||||| :::: ||||| |||||
538 euGlnIleIleGluInAspGlySerGluCysTyrValPheArdLys 554
    :::: ||||| ||||| ||||| :::: ||||| |||||
75216 TCCAGATAATCCAAGAGATAAAGGTTTCAGATTCTTACGTATTTCTGTA 75167
    :::: ||||| ||||| ||||| :::: ||||| |||||
555 TrpGlyArgValGlySerGluLysIleGlyGlyGlnLysLeuGluMe 571
    :::: ||||| ||||| ||||| :::: ||||| |||||
75166 TGGGCCCGAGTTGGAAATGAAAAGATTGGTGTAAACAAAGTGGAGGAA 75117
    :::: ||||| ||||| ||||| :::: ||||| |||||
571 tSerLysThrGluAlaIleLysGluPheLysArgLeuPheLeuGluLys 588
    :::: ||||| ||||| ||||| :::: ||||| |||||
75116 GTCAAAAGTCTGATCGGTTACGAATTCAAACGCTTATTTCTTCAAAAA 75067
    :::: ||||| ||||| ||||| :::: ||||| |||||
588 hrGlyAsnSerTrpGluAlaTrpGluCysLysThrAsnPheArdLysGln 604
    :::: ||||| ||||| ||||| :::: ||||| |||||
75066 CCGGAACACATGGGAATCTTTGGGAACAAAACGAATTTCCAGAAACAA 75017
    :::: ||||| ||||| ||||| :::: ||||| |||||
605 ProGlyArgPheTyrProLeuAspVal..... 613
    :::: ||||| ||||| ||||| :::: ||||| |||||
75016 CCGTGAATAATTCTCCCGTTGCACATTTGAAGCATCCATATGACTATATC 74967
    :::: ||||| ||||| ||||| :::: ||||| |||||
613 ..... 613
    :::: ||||| ||||| ||||| :::: ||||| |||||
74966 TTTGTCTCTTAGATTCCTTATGATGCTTCAATAAACTCTTTGTATC 74917
    :::: ||||| ||||| ||||| :::: ||||| |||||
614 .....AspTyrGlyValLysLysAlaProLysArgLysAspIle 626
    :::: ||||| ||||| ||||| :::: ||||| |||||
74916 GTTCTGTTTAGGATTTAGAGTTAATAAGCAAGTAGCCAAAAAAGAGCCA 74867
    :::: ||||| ||||| ||||| :::: ||||| |||||
627 SerGluMetLysSerSerLeuAlaProGlnLeuLeuGluMetLysMe 643
    :::: ||||| ||||| ||||| :::: ||||| |||||
74866 TTTTCAGACAGTAGCAACCTTGCTCCATCATTAATAGAAATTTGATCAAGAT 74817
    :::: ||||| ||||| ||||| :::: ||||| |||||
643 tLeuPheAsnValGluThrTyrArg..Ala..... 652
    :::: ||||| ||||| ||||| :::: ||||| |||||
74816 GCTTTTGTATGTTGGAACCTTACAGGTTAGAACTTATTTGGCATTTGCTT 74767
    :::: ||||| ||||| ||||| :::: ||||| |||||
652 ..... 652
    :::: ||||| ||||| ||||| :::: ||||| |||||
74766 CTTCTCCTCTCTCTCCCTAGAACTGTATGAAGTAACCTAATAATTACT 74717
    :::: ||||| ||||| ||||| :::: ||||| |||||
653 .....AlaMetMetGluPheGlu***AsnMetSerGluMetPro 665
    :::: ||||| ||||| ||||| :::: ||||| |||||
74716 TCCGAGATCTGCAATGATGGATTTCGAGATAAATATGTCAGAGATGCCA 74667
    :::: ||||| ||||| ||||| :::: ||||| |||||
666 LeuGlyLysLeuSer*****AsnIleGlu.*** 676
    :::: ||||| ||||| ||||| :::: ||||| |||||
74666 CTTGGGAAGCTCAGAACAACTAATATACGAAGGTTAACTCTACTTCTC 74617
    :::: ||||| ||||| ||||| :::: ||||| |||||
676 ..... 676
    :::: ||||| ||||| ||||| :::: ||||| |||||

```

[illegible]

Mon Mar 11 09:58:37 2002

6 LysAlaThrPheLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLys 22
114 AAGCTCTATCGAGTCAGTACGCAAGAGCGAGCGCGCTCTTGCAGAA 163
22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetVal 39
164 ATGACGGAGGAGATCCCAAGGACTCGCTCCGGATGGCCATCATGGTGC 213
39 InAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
214 AGTCGCCCATGTGTGATGGAAGATCCACACTGGTACCACCTTCTCTGC 263
56 IlePheSerLysLysAsnGlnIleLysSerValAsp... AspValGlu 71
264 TTCTGGAAGTGGGCCCATCCATCCGCCACCCCTGACGTTGAGGTGATG 313
71 ylleAspAlaLeuArgTrpAspGlnGluLysIleArgAsnTyrVal. 87
314 GTTCTCTGAGCTTCGGTGGATGACACAGCAGAAAGTCAAGAACACAGCG 363
88GlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
364 AAGCTGGAGGAGTCACAGCAAGGCCAGGATGGAATTGGTAGCAAGGCA 413
102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
414 GAGAAGACTCTGGGTGACTTTGCACAGCAGATGATGCCAAGTCCACAGAAG 463
114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLys 131
464 TACGTGAAGGGGTGTATGGAGAAGATAGAAAGGGCCAGGTGGCCCTGT 513
131 erAlaLysLeu...GluSerGluGlyProLys...Glylle.....Pro 143
514 CCAAGAAGATGTTGGACCGGAGAGGACACAGCTAGCATGATTGACCGC 563
144 TrpTyrHisAlaAsnCysPhePheGlu.....Valse 154
564 TGGTACCATTCAGCGCTGTTTGTCAACAACAGGAGGAGTGGTTCGG 613
154 rProSerAlaThrValGluLysPheSerGlyTrpAspThrLeuSerAsp 171
614 GCCCGAGTACAGTGGAGTACAGTCAAGGGCTTCAGCTCTCTGCTACAG 663
171 LuAspLysArgThrMetLeuAspLeuValLysLysAsp.....Val 184
664 AGGATAAA.....GAAGCCCTGAAGAAGCAGCTCCAGGAGTC 701
185 GlyAsnAsnGluInAsnLysGlySerLys..... 194
702 AAGAGTGAAGGAAGAAAGAGAAAGGCGATAGGTGGATGGATGAAGT 751
195ArgLysLysSerGluAsnAspIleAspSerTyrLysSerAlaArg 210
752 GCGCAAGAGAAGAAATCTAAAGAAAGAAAGAC.....AAGATAGTAGC 795
210 euAspGluSerThrSerGluGlyThrValArgAsnLysGlycInLeuVal 226
796 TTGAAAAA..... 803
227 AspProArgGlySerAsnThrSerSerAlaAspIleGlnLeuLysLeu 243
804CGCTAAA 811
243 sGluInSerAspThrLeuTrpLysLeuLysAspGlyLeuLysThrHisv 260
812 GCTCAGACAGCAGCTGATCTGGAAGATCAAGCAGCAGCTAAAGAGTGT 861
260 aIserAlaAlaGluLeuArgAspMetLeuGluAlaAsnGlyGlnAspThr 276
862 GTTCAACTAATGACCTGAAGAGTACTCATCTTCAACAGCAGCAAGTG 911
277 SerGlyProGluArgHisLeuLeuAspArgCysAlaAspGlyMetLeuPh 293

912 CTTCTGGGAGTCGGCGATCTTGGACCGAGTAGCTGATGCATGGTGT 961
293 eGlyAlaLeuGlyProCysProValCysAlaAsnGlyMetTyrTyrTyr 310
962 CGGTGCCCTCTTCCTCGCAGGAATGCTCGGTGAGCTGCTTCAAGA 1011
310 snGlyGlnTyrGlnCysSerGlyAsnValSerGluTrpSerLysCysThr 326
1012 GCGATGCCCTATTACTGCACTGGGAGCTGCTGCTGGACCAAGTGTATG 1061
327 TyrSerAlaThrGluProValArgValLysLysLysTrpGlnIleProHi 343
1062 GTCAAGACACACACACCAACCGG.....AAGGAGTGGTAACCCCA... 1103
343 sGlyThrLysAsnAspTyrLeuMetLysTrpPheLysSerGlnLysValL 360
1104AAGGAATCCGAGAAATCTTACCTCAAGAAATGAAGGTTA 1146
360 yLysProGluArgValLeuProMetSer..... 370
1147 AAAGCAGACCGTATATTCCTCCAGAACACCGCCTCCGTGGCCAC 1196
371ProGluLysSerGlySerLysAlaThrGlnArgThrSerLeu 385
1197 CTTCCGCCCTCCACAGCCTCGGCTCTGCTGCTGAACCTCTCTCTC 1246
385 uSerSerLysGlyLeuAspLysLeuArgPheSerValValGly..... 399
1247 AGCAGATAAGCCATTATCCCAAGATGAAGATCTGCTCTCGGGAAGCTGT 1296
400 ..GlnSerLysGluAlaAlaAsnGluTrpIleLysLeu.....Lys 413
1297 CCGGAACAGGATGAAGTCAAGGCCATGATTGAGAACTCGGGGGAAG 1346
414 LeuAlaGlyAlaAsnPheTyrAlaArgValValLysAspIleAspCys 430
1347 TTAACGGGACGCCCAACAGGCTTCCTG.....TGCAT 1381
430 uIleAlaCysGlyGluLeuAspAsnGluAsnAlaGluValArgLysAla 447
1382 CAGCACCACAAAGGAGGTGGAAAGATGAATAAGAGATGGAGGAAGTAA 1431
447 rGArgLeuLysIleProIleValArgGluGlyTyrIleGlyGluCysVal 463
1432 AGAAGCCACATCCGAGTGTCTCTGAGGACTTCTCCAGGACGCTCC 1481
464 LysArgThrLys.....CysCysHisLeuLysIle 474
1482 GCCTCCACCAAGAGCCTTCAGGAGTGTCTTAGCGCACATCTTGCC... 1529
474 eAsnTrpAsnAlaLeuGluSerSerLysGly***ThrValThrValLysv 491
1530 .CCTTGGGGGCGAGGTGAAGCAGCAGCCTGTGAAGTT...GTGCCC 1575
491 allLysGlyArgSerAlaCysSer***SerPro***ValCysLysAsnThr 507
1576 CAAGAGGAAGTCAAGGCTCGCTCTCCAAA.....AAAGCAAG 1616
508 AlaHisIlePro****Trp.....GluLysHisIleGlnCys 520
1617 GCCCAGGTCAAGGAGGAGGTATCAACAAATCTCAAAAGAGAATGAAT 1666
520 s***LeuLys..... 523
1667 AACTCTTAAGAGGAGCAGCAGCTGTGGATCTGTGACTGGAACT 1716
524HisValLeu.....Thr***His 529
1717 CTGCGCATGCTCTGGAGAAAGGTGGAGGTCTTCAGTGCACCTTGGC 1766
530 ***Val.....CysThrGlyTyrValLeuGlnIleI 541
541

```

1767 CTGGTGGACATCGTTAAAGGAACCAACTCCCTACTACAAGCTGCAGCTTCT 1816
541 egluGlnAspGlySerGluCysTyrValPheArgLysTrpGlyArgV 558
      |||:|||||
1817 GGAGGACACAGGAAACAGGTATTGGATATTCAGGTCCTGGGGCCGTG 1866
558 alGlySerGluLysIleGlyGlnLysLeuGluMet...SerLys 573
      |||:|||||
1867 TGGGTACG...GTGATCGGTAGCAACAACCTGGAACAGATGCCGTCCAAG 1913
574 ThrGluAlaIleLysGluPheLysArgLysPheLeuGluLysThrGlyAs 590
      |||:|||||
1914 GAGATGCCATGAGCTACCTCATCAATATATATGAAGAAAAACCGGAA 1963
590 nSerTrpGluAlaTrpGluCysLysThrAsnPheArgLysGlnProGlyA 607
      |||:|||||
1964 CGCTTGGCACTCC.....AAAATTTCCACCAAGTATCCCAAAA 2001
607 rgPheTyrProLeuAspValAspTyrGlyVal...LysLysAlaProLys 622
      |||:|||||
2002 AGTTCTACCCCTGGAGATTGACTATGGCCAGGATGAAGGAGCAGTGAAG 2051
623 ArgLysAspIleSerGlu...MetLysSerSerLeuAlaProGlnLeuLe 638
      |||:|||||
2052 AGCTGACAGTAAATCTCGCACCAAGTCCAAGCTCCCAAGCCAGTTCA 2101
638 uGluLeuMetLysMetLeuPheAsnValGluThrTyrArgAlaAlaMetM 655
      |||:|||||
2102 GGACCTCATCAAGATGATCTTTGATGTGGAAGTATGAAGAAGCCATGG 2151
655 etGluPheGlu***AsnMetSerGluMetProLeuGlyLysLysSer*** 671
2152 TGGAGTATGAGATCGACCTTTCAGAAGATGCCCTTGGGAAGCTGAGCAAA 2201
672 ***AsnIleGlu***GlyPheGluAlaLeuThr*****LeuPh 688
2202 AGCAGATCAGCGCCGATCTCATCCTCAGT..... 2234
588 eGluGlyHisArg***SerSerThrGlyLeu***GluLysAla***LeuL 705
      |||:|||||
2235 .GAGGTCCAGCAGCGCGGTGTCTCAGGCAGCAGCAGCTCTCAGATCCTGG 2283
705 eu*****PheSerLeuLeuSerLeuLeuPheIleLeuIle 721
      |||:|||||
2284 ATCTCTCAATCCTTTTACACCTCATCCTCCCAAGCTTT..... 2324
722 LeuTyrGlyMetArgMet.....IleSerTyr 730
2325 .....GGGATGAAGAAGCCTCGCTCCTGAACAATGCAGACAGTGC 2368
730 rSerLysAlaLysMetLeuGluAlaLeuGlnAspIleGluIleAlaSerL 747
      |||:|||||
2369 GGCCAAAGTGGAAATGCTTGCACAACTGCTGGACATCGAGGTGGCTACA 2418
747 ysIleVal.....GlyPheAspSerAspSerAspGluSerLeuAspasp 761
      |||:|||||
2419 GTCTGCTCAGGGAGGCTGTGATGATAGCAGCAAGATCCCATCGATGTC 2468
762 LysTyrMetLysLysCysAspIleThrProLeuAlaHisAspSerG1 778
      |||:|||||
2469 AACTATGAGAAGCTCAAACTGACATTAAGTGGTTGACAGAGATTCGA 2518
778 uAspTyrLysLeuIleGluGlnTyrLeuLeuAsnThrHisAlaProThrH 795
      |||:|||||
2519 AGAGCCGAGATCATCAAGAGATGTTTAAAGAACACTCATGCAACACAC 2568
795 isLysAspTrpSerLeuGluLeuGluValPheSerLeuAspArgAsp 811
      |||:|||||
2569 ACAATGCGTATGACTTGGAGTCAATGATATCTTTAAGATAGAGCGTGA 2618
812 GlyGluLeuAsnLysTyrSerArgTyrLysAsnLeuLeuHisAsnLysMe 828
      |||:|||||
2619 GCGCAATGCCAGCGTTACAGACCCCTTTAAG...CAGCTTCAACCCGAAG 2665
828 tLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyIleLeuSerG 845
      |||:|||||
2666 ATTGCTGTGGCAGCGGTCCAGGACCACTTGTGGGATCCTGTCC 2715
845 lnglyLeuArgIleAlaProGluAlaProValThrGlyTyrMetPhe 861
      |||:|||||
2716 AGGCTCTTCGGATAGCCCGCTGAAGCCCGCTGACAGGCTACATGTTT 2765
862 GlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCy 878
      |||:|||||
2766 GGTAAAGGATCATTTCTGCTGACATGCTCCAAGAGTGCACCACTACTG 2815
878 styValAspArgAsnAsnProValGlyLeuMetLeuLeuSerGluValA 895
      |||:|||||
2816 CCATAGCTCTCAGGGAGACCCCAATAGGCTTAATCTGTGGGAGAGTTG 2865
895 laLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetAspLysPro 911
      |||:|||||
2866 CCCTTGGAAACATGTATGAATGAAGCAGCGCTTCACATATCAGCAAGT 2915
912 ProArgGlyLysHisSerThrLysGlyLeuGlyLysThrValProLeuG1 928
      |||:|||||
2916 CCCAAGGCAAGCAGTGTCAAGGTTTGGCAAACTACCCCTGATCC 2965
928 uSerGluPheValLysTyrArgAspAspValValValProCysGlyLysP 945
      |||:|||||
2966 TTCAGCTAACATTAAGT...CTGGATGGGTAGACGCTCTCTTGGGACCG 3012
945 roValProSerSerIleArgSerSerGluLeuMetTyrAsnGluTyrIle 961
      |||:|||||
3013 GGATTTTCATCTGGTGTGAATGACACCTCTCTACTATATACAGAGTACAT 3062
962 ValTyrAsnThrSerGlnValLysMetGlnPheLeuLysValArgph 978
      |||:|||||
3063 CTCTATGATATTGCTCAGGTAAATCTGAAGTATCTGCTGAAACTGAAATT 3112
978 ehIshisLys 981
      |||:|||||
3113 CAATTTTAAG 3122
seq_name: gb_pr:HUMPOLP
seq_documentation_block:
LOCUS HUMPOLP 3640 bp mRNA PRI 08-JAN-1995
DEFINITION Human poly(ADP-ribose) polymerase mRNA, complete cds.
ACCESSION M18112
VERSION M18112.1 GI:190166
KEYWORDS polymerase.
SOURCE Human SV40 transformed fibroblast, cDNA to mRNA, clone pPAP.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Uchida,K., Morita,T., Sato,T., Ogura,T., Yamashita,R., Noguchi,S.,
Suzuki,H., Nyunoya,H., Miwa,M. and Sugimura,T.
TITLE Nucleotide sequence of a full-length cDNA for human fibroblast
poly(ADP-ribose) polymerase
JOURNAL Biochem. Biophys. Res. Commun. 148 (2), 617-622 (1987)
MEDLINE 88076933
COMMENT Draft entry and computer readable sequence for [1] kindly provided
by K.Uchida, 02-MAR-1988.
FEATURES
source
1..3640
/db_xref="taxon:9606"
/map="1441-q42"
<1..3640
/note="POLP mRNA"
140..3184
/ gene="ppol"
140..3184
/ gene="PPOL"

```


2279 AGT.....GAGTCCAGCAGCGCGTGTCTCAGGCGAG 2311

698 u**GluLysAla***LeuLeu*****PheSerLeuLeuS 715
||||:||||| ||| ||| :|||:||||:|
2311 CAGCGACTCTCAGATCCTGGATCTCTCAAATCGCTTTTACACCCGTGATCC 2360

715 erLeuLeupheleleuLeuLeuTyrrGlyMetArgMet..... 727
||| ||| |||:||||:|
2361 CCCACGACTTT.....GGGATGAAGAAGCCCTCCGCTCCTG 2395

728IleSerTyrrSerLysAlaLysMetLeuGluAlaLeuGlu 740
|||:||||:||||| |||:||||:|
2396 ACAATGCAGACAGTCTCGACGCCNAGTGGAAATGCTTGACACCHGCT 2445

740 nAspIleGluIleAlaSerLysIleVal.....GlyPheAspSerAsps 755
||||:||||:||||| ||| ||| :|||:||||:|
2446 GGACACTCAGGTGGCGCTACAGTCTGCTCAGGGGAGGGTCTCATGATAGCA 2495

755 erAspGluSerLeuAspAspLysTyrrMetLysLeuHisCysAspIleThr 771
|| ||| :|||:||||| ||| ||| :|||:||||:|
2496 GCAAGATGCCATCATGATCAACTATGAGAAGCTCAAACTGACATTAAAG 2545

772 ProLeuAlaHisAspSerGluAspTyrrLysLeuIleGluGlnTyrrLeu 788
:||| :|||:||||:||||| ||| ||| :|||:||||:|
2546 GTGGTTGACAGATTCTGAAGAAGCCGAGATCATCAGGAATATGTTAA 2595

788 uAsnThrHisAlaProThrHisLysAspTrpSerLeuGluLeuGluGluV 805
||||:||||:||||| ||| ||| :|||:||||:|
2596 GAACACTCATGCACACACACAATGCGCTATGACTTCGGAATCATCGATA 2645

805 alPheSerLeuAspArgAspGlyGluLeuAsnLysTyrrSerArgTyrrLys 821
:||| :|||:||||:||||| ||| ||| :|||:||||:|
2646 TCTTTAAGATAGCGGTGAAGCGAATGCCAGCGTTACAGCCCTTTAAG 2695

822 AsnAsnLeuHisAsnLysMetLeuLeuThrPheHisGlySerArgLeuThrAs 838
:|||:||||:||||| ||| ||| :|||:||||:|
2696 ...CAGCTTCATAACCGAAGATTGCTGTGGCAGCGGTCCAGGACCCACAA 2742

838 nPheValGlyIleLeuSerGlnGlyLeuArgIleAlaProGluAlaLap 855
|||:||||:||||:||||| ||| ||| :|||:||||:|
2743 CTTTCTGGGATCTGTGCCAGGTCCTCGGATAGCCCCGCCCTGAGCGC 2792

855 roValThrGlyTyrrMetPheGlyLysGlyLeuTyrrPheAlaAspLeuVal 871
|||:||||:||||:||||| ||| ||| :|||:||||:|
2793 CCGTGACAGGCTACATGTTTGGTAAGGGATCTATTTTCGTGCATGGTC 2842

872 SerLysSerAlaGlnTyrrCysTyrrValAspArgAsnAsnProValGlyLe 888
:|||:||||:||||:||||| ||| ||| :|||:||||:|
2843 TCCAAGATGCCACTACTGCGCATCTGCTCTCAGGGAGACCCCAATAGGCTT 2892

888 uMetLeuLeuSerGluValAlaLeuGlyAspMetTyrrGluLeuLysLysL 905
:|||:||||:||||:||||| ||| ||| :|||:||||:|
2893 AATCCTGTTGGAGAGTTGCCCTTGGAAACATCTATGAATCAAGCAGC 2942

905 laThrSerMetAspLysProProArgGlyLysHisSerThrLysGlyLeu 921
:|||:||||:||||:||||| ||| ||| :|||:||||:|
2943 CTTTCATATCAGCAAGTTATCCCAAGGGCAGGACAGTGTCAAGGTTTG 2992

922 GlyLysThrValProLeuGluSerGluPheValLysTrpArgAspAspVa 938
:|||:||||:||||| ||| :|||:||||:|
2993 GGCAAACTACCCCTGATCCTTCAGCTAAACATTAGT...CTGATCGTGT 3039

938 lValValProCysGlyLysProValProSerSerIleArgSerSerGluL 955
|| ||||| ||| ||| :|||:||||:|
3040 AGAGCTTCTCTTGGGACCGGGATTTCATCTGCTGTGTGAATGACACCTCTC 3089

955 euMetTyrrAsnGluTyrrIleValTyrrAsnThrSerGlnValLysMetGln 971
:|||:||||:||||:||||| ||| ||| :|||:||||:|
3090 TACTATATACCGATACATGTCTATGATATTGCTCAGGTAAATCTGAAG 3139

972 PheLeuLeuLysValArgPheHisHisLys 981
:|||:||||:||||:||||| ||| ||| :|||:||||:|
3140 TATCTGTGAACCTGAATTCAAATTTTAAAG 3169

||||| ||||| ||| ... : : :
870 TCTGCTTCTTTGTGAAGAGTGTAAGGGCAGCTTGTGTTCGAAGACGCATG 919

312 lntYrGlncysSerGlyAsnValSerGluTrpSerLysCysThrTyrrSer 328
||| ||||| ||||| ||||| ||||| : : :
920 CATACTACTGCTCAGGGGACATTACTGCTGACTAAGTGTGTGGCTAAA 969

329 AlaThrGluProValArgValLysLysLysTrpGlnIleProHisGlyTh 345
... ||| ||| ||| ||| |||||
970 ACACAGACTCCCAANTAGG.....AAGACTGGGTAAATCCCG..... 1005

345 rLysAsnAspTryLeuMetLysTrpPhelySerGlnLysValLysLysP 362
||||: : : : : : : : : : : : :
1006 .AAGGAATTCGTGAGATTCTTATCTCAAATAATTTAAGTGAAGAAGC 1054

362 roGluArgValLeuProPrometSerpro 371
||| ||| ||| ||| ||| : : :
1055 AGGATAGGATATTCCTCTGAGGCTGCACTGTGAACCTGTCACCACCAT 1104

372 GluLysSerGlySerLysAlaThrGlnArgThrSerLeuLeuSerSerLy 388
||| ||| : : : : : : : : : : : : :
1105 CCACCTGCACTCCGCTCCATTGACAGAGACTGCACTGCACCCCACAGACA 1154

388 sGlyLeuaspLysLeuArgPheSerValValglyGlnSerL 402
||| ||| : : : : : : : : : : : : :
1155 ACCACTGCAACATGAAATCCTGACTCTTTGGAAAGCTGTCCAAGAACA 1204

402 ySGluAlaAlaAsnGluTrpIleGluLysLeuLysLeuAlaGlyAlaAsn 418
||||| ||| ||| ||| ||| : : : : : : :
1205 AGGAGGAGTGAAGAAATATGTGGGAGAGCTG.....GGAGGAAG 1245

419 PheTYrAlaArgValValLysAspileAspCysLeuileAlacysGlyGI 435
||| ||| : : : : : : : : : : : : :
1246 ATGCAACAACAGCTAACAGGCCACCTGTGCATCAGCACACAAAGGA 1295

435 uLeuAspAsnGluAsnAlaGluValArgLysAlaArgArgLeuLysIleP 452
||| ||| : : : : : : : : : : : : :
1296 GGTGGAAAAATGACCAAGAAGATGGGAAGTGAAGATGCCAAAGTTC 1345

452 roIleValArgGluGlyTrpIleGlyGluCysVallysArgThrLys... 467
||| ||| : : : : : : : : : : : : :
1346 GTGTGCTCTCAGAGGATTTCTTAAAGATGTGAATCTTCCCAACAGGCG 1395

468CysCysHisLeuileCysileAsnTrpAsnAlaLe 479
||| ||| : : : : : : : : : : : : :
1396 TTTCAGGAGCTTCTCTCTCCATCAATTCA...CCTGGGGTGCACA 1442

479 uGluSerSerLysGly***ThrValThrVallysValLysGlyArgSera 496
: : : : : : : : : : : : : : : : :
1443 GGTGAAAACGGAGCACCAGGAGTGGCGGTGGATGGGAAG..... 1482

496 IaCysSer***SerPro***ValCysLysAsnThrAlaHisIle..... 510
: : : : : : : : : : : : : : : : :
1483 ..TCGAGC....AAGCTGCAAAATATGAAGAGTGTGGGAGGTCCAAGAA 1527

511Pro*****TrpGluLysHisIleGlnCys***LeuLys... 523
||||| ||||| : : : : : : : : : : : : :
1528 GAACAAGGACCTAGCAAGTCTGAGAAGAAATGAAGCTAACACTTAAGG 1577

524HisValL 526
: : : : : : : : : : : : : : : : :
1578 TGGAGCAGCGTAGATCCTGATTCTGGCTTGGAGGATTCGTCTCATGTCT 1627

526 eu.....Thr***His***Val..... 531
: : : : : : : : : : : : : : : : :
1628 TTGAAAAGGTGGGAAGATTTTCAGTGCACACTTGGCGCTAGTGATATT 1677

532CysThrGlyTYrTYrValLeuGlnIleIleGluGlnAspAs 545
: : : : : : : : : : : : : : : : :
1678 GTGAAAGSAACAAATCCTATTACAANAATGCAGCTGCTGGAGGATGCAC 1727

545 pGlySerGluCystyrValpheArgLysTrpGlyArgValGlySerGluL 562
||| : : : : : : : : : : : : : : : : :

[illegible]


```

849 leaIaProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeu 865
|||||
2627 TAGTCCCTCGAAGCTCCTGTGACCGGTACATGTTGGGAGGCATC 2676
|||||
866 TyrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspAr 882
|||||
2677 TACTTTCGAGATATGTTTCCAGAGAGTGCCAACTACTGTACACATCTCA 2726
|||||
882 gAsnAsnProValGlyLeuMetLeuSerGluValAlaLeuGlyAspM 899
|||||
2727 AGCTGTACCAATTTGGGTTAATACTACTGGGAGAGTTGCCCTCGGAATA 2776
|||||
899 etTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLys 915
|||||
2777 TGTATGAGCTTAAGATAGCTTCTACATACAAATATGCCCAAGGAAAA 2826
|||||
916 HisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVa 932
|||||
2827 CATAGTGTGAAGGCTTGGCAAACTGCACCT...GATCCACACAGCCAC 2873
|||||
932 LysTrpArgAspValValValProCysGlyLysProValProSerS 949
|||||
2874 TACAACCTTTCATGCTGTAGAGTTCCCTTAGGGAATGGGATCTCAACAG 2923
|||||
949 erIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThr 965
|||||
2924 GAATTAATGATACCTGCTCTGTACAAATGAATATATGTTGATGATGTT 2973
|||||
966 SerGlnValLysMetGlnPheLeuLeuLysValArgPheHisLys 981
|||||
2974 GCTCAGTAAATCTGAAGTACCTGCTGAACTGAAATCAACTATAAG 3021
|||||
seq_name: gb_pr:HUMRISDAD
seq_documentation_block:
LOCUS HUMRISDAD 3795 bp mRNA PRI 09-JAN-1995
DEFINITION Human poly(ADP-ribose) synthetase mRNA, complete cds.
ACCESSION J03473
VERSION J03473.1 GI:337423
KEYWORDS poly(ADP-ribose) synthetase.
SOURCE human placenta, cDNA to mRNA (library of H.Okayama), clones pPARS[1,11,21,32,32,41,-F].
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3795)
Kurosaki,T., Ushiro,H., Mitsuuchi,Y., Suzuki,S., Matsuda,M.,
Matsuda,Y., Katunuma,N., Kangawa,K., Matsuo,H., Hirose,T.,
Inayama,S. and Shizuta,Y.
Primary structure of human poly(ADP-ribose) synthetase as deduced
from cDNA sequence
J. Biol. Chem. 262 (33), 15990-15997 (1987)
88058958
Draft entry and printed copy of sequence for [1] kindly provided by
Y.Shizuta, 23-NOV-1987.
FEATURES
source
1..3795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1q41-q42"
<1..3795
/notes="poly(ADP-ribose) synthetase mRNA"
96..3140
/gene="PPOL"
/notes="poly(ADP-ribose) synthetase"
96..3140
/gene="PPOL"
/notes="PPOL"
/codon_start=1
/db_xref="GDB:G00-119-508"
/protein_id="AAB59447.1"
/db_xref="GI:337424"
/translation="MAESSDKLYRVEYAKSERASCKKCSIPKDSLRLMAIMVQSPMF

```

old_sequence

/gene="PPOL"

/note="cg"

/citation=[1]

/replace=""

1193..1199

/gene="PPOL"

/note="ccac"

/citation=[1]

/replace=""

BASE COUNT 1048 a 847 c 1037 g 863 t

ORIGIN 73 bp upstream of AluI site.

alignment_scores:

Quality: 1593.00

Ratio: 2.309

Percent Similarity: 64.426

Length: 1071

Gaps: 37

Percent Identity: 36.321

alignment_block:

US-09-236-995D-2 x HUMRISDAD ..

Align seg 1/1 to: HUMRISDAD from: 1 to: 3795

6 LysAlaTriPysAlaGluTyrAlaLysSerGlyArgAlaSerCysLys 22

|||||

114 AAGCTTATCGAGTCGAGTACGACCAAGAGCGAGCGGCTCTTGGCAAGAA 163

|||||

22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetVal 39

|||||

164 ATCAGCGGAGAGCATCCCAAGAGCTCGCTCCGATGGCCATCATGGTGC 213

|||||

39 InAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55

|||||

214 AGTCGCCCATGTTTGTATGTAAGAAAGTCCACACTGGTACCACCTTCTCTGC 263

|||||

56 IlePheSerLysLysAsnGlnIleLysSerValAsp...AspValGlu 71

|||||

264 TTCTGGAAGTGGCCACCTCCATCCGCGCACCTGACGTTGAGGTGATG 313

|||||

71 yIleAspAlaLeuArgTrpAspGlnGluLysIleArgAsnTyrVal.. 87

|||||

314 GTTCTCTGAGCTTCGGTGGGATGACCAAGTCAAGAAAGTCAAGACAGCGG 363

|||||

88GlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101

|||||

364 AAGCTGGAGGAGTGACAGGCAAGGCCAGGATGGATTTGATAGCAAGGCA 413

|||||

102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114

|||||

414 GAGAGACTCTGGGTGACTTTGGCAGAGATATGCCAAGTCCCAAGAGAG 463

|||||

114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLeu 131

|||||

464 TACGTGCAAGGGGTGTATGAGAAGATAGAAAAGGGCCAGGTGCGCTGT 513

|||||

DGKVPWHYHPSFWKVGHSIRHPDVEVDFSELRWDDQOKVKVTAEGAGVCTCKGDDGI
GSAKRTLDGFAAEYAKSNRSTCKGMEKIEKQVRLSKKNVDPEKPOLGMDRWHYP
GCFVKNREELGERPEYSASQAKGSLIATEDEALKQLPGVSEKRRGKQVDPGVE
VAKKSKKEDKSKLEKAKQNDLIWNIKDELLKVCSTNDLKELLIFNKQVPSGE
SATLDRVADGMVFGALLPCEEGSLVFKSDAYCTGDTAVTKCMKVQTQTPNRKEW
TPKEPREISVLKLVKKODRIPPPETSASVAATPPPSASAPAAVNSSASADKPLSN
MKILTLGKLSRNDKDEKAMIEKLGKLTGTANKASLCISTKKEVKMNKMEVEKEAN
IRVSEDFLDQVDSASTSKQELFLAHILSPWGAEVKAEPEVVAAPRGKSGAALS
GOVKEGINKSEKMKLTILGGGAAPDPSGLHSAHVLEKGGKVFSAITGLVDIVKGT
NSYKQLLEDDEKRNRYWIFRSRGRVGTWIGSNKLEQMPKEDAIHFMKLYEETGN
AWHSNFTKYPKFFPLEIDYGODEAVKLTVPNGTKSLPKPQVODLKMIFVESM
KMAVEYEDLOKMPGLKSKROIOAAYSTLSEVOQAVSOGSDSDILOLNSRFYTLI
PHDGMKPPPLNNAVSQAKVEMLNDLIEVAYSLRGGSDSDSKDPIIDVNYEKLK
TDIKVDRDSEEAELIRKYVKNTHAIDLEIVAYSLRGGSDSDSKDPIIDVNYEKLK
RRLHWSRTTFNAGLSQGLRIAPPEPVTFYMGFKGIFADVMVSKNSANYCHTSOGD
PIGLILGEVALGNMVELHASHISKLPKHSVKGIGTKTTPDPSANISLDGVDVPLG
TGTSVGWNTSLLYNEIVYDIAQVNLKYLKLFNFKTSLW"

```

131 erAlaLysLeu...GluSerGluGlyProLys...Glylle.....Pro 143
132  |||||:  ::  |||  |||||:  |||||:
514 CAAAGAGATGGTGGACCGGAGAGCCACAGTAGCATGGATGACCGC 563
144 TrpTyrHisAlaAsnCysPhePheGlu.....Valse 154
145  |||||:  ::  |||||:  ::
564 TGGTACCATCCAGCTGCTTTGTCAAGAACACAGGAGGAGCTGGGTTTCCG 613
154 rProSerAlaThrValGluLysPheSerGlyTyrAspThrLeuSerAspG 171
155  |||||:  ::  |||||:  ::  |||||:  |||||:  |||||:  |||||:
614 GCCGAGTACAGTCGAGTACAGTCAAGGCTTCAGCTCTCTGTCACAG 663
171 luAspLysArgThrMetLeuAspLeuValLysLysAsp.....Val 184
172  |||||:  ::  |||||:  ::  |||||:  |||||:
664 AGGATAAA.....GAAGCCCTGAAGACAGCTCCAGAGGTC 701
185 GlyAsnAsnGluGlnAsnLysGlySerLys..... 194
195  |||||:  ::  |||||:  |||||:  |||||:
702 AAGAGTGAAGGAAGAGAGAAAGCGATAGGTGGATGGATGAAGT 751
195 ....ArgLysLysSerGluAsnAspLysSerTyrLysSerAlaArgL 210
196  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
752 GCGAAGAGAAATCTAAAGAAAGAAAGAC.....AAGGATAGTAAGC 795
210 euAspGluSerThrSerGluGlyThrValArgAsnLysGlyGlnLeuVal 226
211  |||||:  ::  |||||:  ::  |||||:  |||||:
796 TTGAAGAAAGCC..... 806
227 AspProArgGlySerAsnThrSerSerAlaAspIleGlnLeuLysLeu 243
228  |||||:  ::  |||||:  |||||:  |||||:  |||||:
807 .....CTAAA 811
243 sGluGlnSerAspThrLeuTrpLysLeuLysAspGlyLeuLysThrHis 260
244  |||||:  ::  |||||:  |||||:  |||||:  |||||:
812 GCCTCAGACGACCTGATCTGGAACTCAAGGACGAGCTAAAGAGTGT 861
260 alSerAlaLagLeuArgAspMetLeuGluAlaAsnGlyGlnAspThr 276
261  |||||:  ::  |||||:  |||||:  |||||:  |||||:
862 GTTCAACTAATGACCTGAAGGAGCTACTCATCTCAACAAGCAGCAAGT 911
277 SerGlyProGluArgHisLeuLeuAspArgCysAlaAspGlyMetLeuPh 293
278  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
912 CCTTCTGGGGAGCTCGCGCATCTTGGACCGAGTAGCTGATGCTGTGT 961
293 eGlyAlaLeuGlyProCysProValCysAlaAsnGlyMetTyrTyr 310
294  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
962 CGGTGCTCTCTTCCCTCGGAGGATGCTCGGGTCAGCTGGTCTCAAGA 1011
310 snGlyGlnTyrGlnCysSerGlyAsnValSerGluTrpSerLysCysThr 326
311  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1012 GCGATGCTTATTACTGCATGGGGAGCTCCTCGGACCAAGTGTATG 1061
327 TyrSerAlaThrGluProValArgValLysLysLysTrpGlnIleProHi 343
328  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1062 GTCAAGACACACACACCCACCGG.....AAGGAGTGGTAAACCCCA... 1103
343 sGlyThrLysAsnAspTyrLeuMetLysTrpPheLysSerGlnLysVal 360
344  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1104 .....AAGGAATCCGAGAATCTCTTACCTCAAGAAATGAAGTTA 1146
360 yLysProGluArgValLeuProMetSer..... 370
361  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1147 AAAAGCAGGACCGTATATTCCCGCCAGAAACCCAGCGCTCCGCGGCC 1196
371 .....ProGluLysSerGlySerLysAlaThrGlnArgThrSerLe 384
372  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1197 AGCCTCCGCTCCACAGCTCGGCTCTGCTGTGTAACCTCCTCTGC 1246
384 uLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValGly... 399
395  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1247 TTCAGCAGATAAGCATATCAACATGAAGTCTGTACTCTCGGGAAGC 1296
400 .....GlnSerLysGluAlaAlaAsnGluTrpIleGluLysLeu... 412

```

```

1297 TGTCCCGGAACAGGATGAAGTGAAGGCCATGATTGAGAACTCGGGGG 1346
1298  |||||:  ::  |||||:  |||||:  |||||:  |||||:
413 LysLeuAlaGlyAlaAsnPheTyrAlaArgValValLysAspIleAspCy 429
1340  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1347 AAGTTAACGGGACGCCCAACAGGCTTCCCTG.....TG 1381
429 sleulleAlaCysGlyGluLeuAspAsnGluAlaGluValArgLysA 446
1382  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1382 CATCAGCACCAAAAAAGAGGCTGGAAGATGAATAAGAGATGGAGGAAG 1431
446 laArgArgLeuLysIleProIleValArgGluGlyTyrIleGlyLys 462
1432  |||||:  ::  |||||:  |||||:  |||||:  |||||:  |||||:
1432 TAAAGAGACCAACATCCAGTGTGTCTGAGGACTTCTCCAGGACGTC 1481
463 ValLysArgThrLys.....CysCysHisLeuIleCy 473
1482  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1482 TCCGCTCCACCAAGAGCCTTCAGGAGTTGTTCTTAGCGCATCTTGTG 1531
473 sIleAsnTrpAsnAlaLeuGluSerSerLysGly***ThrValThrVal 490
1532  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1532 C...CCTTGGGGGCGAGAGGTGAAGGCAGAGCCTGTTGAAGTT...GTGG 1575
490 ysValLysGlyArgSerAlaCysSer***SerPro***ValCysLysasn 506
1576  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1576 CCCCAGAGGGAAGTCAGGGGCTGCGCTCCAAA.....AAAAGC 1616
507 ThrAlaHisIlePro****Trp.....GluLysHisIleGl 519
1617  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1617 AAGGCCAGGTCAAGGAGGAGGTATCAACAAATCTGAAAAGAGAGTCAA 1666
519 nCys***LeuLys..... 523
1667  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1667 ATTAACCTCTTAAAGGAGGAGCAGCTGTGGATCTCTGACTGGAAC 1716
524 .....HisValLeu.....Thr*** 528
1717  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1717 ACTGCGCCATGTCTGGAGAAAGGTGGAGAGGTCTTCAGTGGCCACCTT 1766
529 His***Val.....CysThrGlyTyrTyrValLeuGlnI 540
1767  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1767 GGCCTGTGGAGCATCTGTTAAAGAACCAACTCTCTACAGCTGCAGCT 1816
540 eIleGluAlaAspAspGlySerGluCysTyrValPheArgLysTrpGly 557
1817  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1817 TCTGGAGGACGACAAAGAAACAGGTATTGGATATTCAGTCTCTGGGCC 1866
557 rgValGlySerGluLysIleGlyGlnLysLeuGluMet...Ser 572
1867  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1867 GTGTGGGTAGC...GTGTCGTAGTACCAAACTGGAGACAGATGCCGTCC 1913
573 LysThrGluAlaIleLysGluPheLysArgLeuPheLeuLysThrGl 589
1914  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1914 AAGGAGTGCATGAGTGCATCTCATCAATATATGAAGAAAAACCGG 1963
589 yAsnSerTrpGluAlaTrpGluCysLysThrAsnPheArgLysGlnPro 606
1964  |||||:  ::  |||||:  |||||:  |||||:  |||||:
1964 GAAGCTTGGCCTCC.....AAAAATTCACGAAGTATCCCA 2001
606 lyArgPheTrpProLeuAspValAspTyrGlyVal...LysLysAlaPro 621
2002  |||||:  ::  |||||:  |||||:  |||||:  |||||:
2002 AAAAGTTCTACCCCTCGAGTGCATCTATGCCAGGATGAAGAGCAGTC 2051
622 LysArgLysAspIleSerGlu...MetLysSerSerLeuAlaProGlnLe 637
2052  |||||:  ::  |||||:  |||||:  |||||:  |||||:
2052 AAGAAGTGCAGTAAATCTCTGCGACCAAGTCCCAAGCTCCCAAGCAGT 2101
637 uLeuGluLeuMetLysMetLeuPheAsnValGluThrTyrArgAlaAla 654
2102  |||||:  ::  |||||:  |||||:  |||||:  |||||:
2102 TCAGGACCTCATCAAGATGATCTTTGATGTGAAGATGTAAGAAAGCCA 2151
654 etMetGluPheGlu***AsnMetSerGluMetProLeuGlyLysLeuSer 670
|||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:

```

```
2152 TGGTGGAGTATGAGATCGACCTTCAGAAAGATGCCCTTGGGGAAGCTGAGC 2201
671 *****AsnIleGlu***GlyPheGluAlaLeuThr*****Le 687
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2202 AAAGGCGAGATCCAGGCGGATACCTCCTCCTCAGT..... 2237
687 uPheGluGlyHisArg***SerSerThrGlyLeu***GluLysAla***L 704
|||:::|||||:::|||||:::|||||:::|||||:::
2238 ...GAGGTCAGCAGCGGTGCTGCTGACAGCAGCAGCAGCTCAGATCC 2283
704 euLeu*****pHeSerLeuLeuSerLeuLeuPheIleLeu 720
|||:::|||||:::|||||:::|||||:::|||||:::
2284 TGGATCTCTCAATCGCTTTTACACCTGATGCCGCCACGACTTT..... 2327
721 IleLeuTyrGlyMetArgMet.....IleSe 729
|||||:::
2328 .....GGGATGAGAGAGCCTCCCTCCTGAACAATGCAGACAGTGT 2368
729 rTyrSerLysAlaLysMetLeuGluAlaLeuGlnAspIleGluLeAlas 746
:::|||||:::|||||:::|||||:::|||||:::
2369 GCAGGCCAAGGTGAAATGCTTGACAACCTGCTGGACATCGAGGTGGCCT 2418
746 erLysIleVal.....GlyPheAspSerAspSerGluSerLeuAsp 760
:::|||||:::|||||:::|||||:::|||||:::
2419 ACAGTCTGCTCAGGGGAGGCTGCTGATAGCAGCAAGGATCCCATCGAT 2468
761 AspLysTyrMetLysLeuHisCysAspIleThrProLeuAlaHisAspSe 777
:::|||||:::|||||:::|||||:::|||||:::
2469 GTCACATATGAGAGCTCAAAAGTGCATTAAGTGGTTACAGAGATTC 2518
777 rGluAspTyrLysLeuIleGluInTyrLeuLeuAsnThrHisAlaProT 794
|||||:::|||||:::|||||:::|||||:::
2519 TGAAGAAGCCGAGATCATCAGGAAGTATGTTAAGAACACTCATCAACCA 2568
794 hrHisLysAspTyrSerLeuGluLeuGluAlaPheSerLeuAspArg 810
|||||:::|||||:::|||||:::|||||:::
2569 CACCAATGCGTATGACTTGGAACTGATGATATCTTTAAGATAGACGT 2618
811 AspGlyGluLeuAsnLysTyrSerArgTyrLysAsnAsnLeuHisAsnLy 827
:::|||||:::|||||:::|||||:::|||||:::
2619 GAAGCGAATGCCAGGTTACAGCCCTTTAAG...CAGCTTCATAACCG 2665
827 smetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyIleLeu 844
:::|||||:::|||||:::|||||:::|||||:::
2666 AAGATTGCTGTGGCAGCGGTCAGGACCACTTTGCTGGGATCCTGT 2715
844 erGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyTyrMet 860
|||||:::|||||:::|||||:::|||||:::
2716 CCCAGGCTCTTCGATAGCCCCGCTGAAGCGCCGTCAGACGGGTACATG 2765
861 PheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTy 877
|||||:::|||||:::|||||:::|||||:::
2766 TTTGGTAAGGATCTATTTCGCTGACATGCTCCAAAGAGTGCACACTA 2815
877 rCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSerGlu 894
|||||:::|||||:::|||||:::|||||:::
2816 CTGCATACGCTCAGGGAGACCAATAGGCTTAATCCTGTGGGAGAAG 2865
894 alAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetAspLys 910
|||||:::|||||:::|||||:::|||||:::
2866 TTGCCCTTGGAAACATGATGAACTGAAGCAGCGCTTCACATATCAGCAAG 2915
911 ProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrValProLe 927
|||||:::|||||:::|||||:::|||||:::
2916 TTACCAGGCGCAGCAGCTGCTCAAGGTTTGGGCAAACTACCCCTGA 2965
927 uGluSerGluPheValThrArgAspValValValProCysGlyL 944
|||:::|||||:::|||||:::|||||:::
2966 TCCCTTACCTAACATTAGT...CTGGATGGGTAGACCTTCTCTTGGGA 3012
944 ysProValProSerSerIleArgSerSerGluLeuMetTyrAsnGluTyr 960
:::|||||:::|||||:::|||||:::|||||:::
3013 CCGGGATTTTCACTCTGCTGTAATGACACCTCTCTACTATATAAGAGTAC 3062
```

```
961 IleValTyrAsnThrSerGlnVallLysMetGlnPheLeuLysValar 977
|||||:::|||||:::|||||:::|||||:::|||||:::
3063 ATTGCTATGATATTGCTCAGCTAAATCTGAAGTATCTGCTGAAACTGAA 3112
```

```
977 gPheHisHisLys 981
:|||||:::|||||
3113 ATTCAATTTTAAG 3125
```

seq_name: gb_ro:BC012041

seq_documentation_block: 3845 bp mRNA ROD 02-AUG-2001
LOCUS BC012041
DEFINITION Mus musculus, Similar to ADP-ribosyltransferase (NAD+); poly
(ADP-ribose) polymerase), clone MGC:6498 IMAGE:2648390, mRNA,
complete cds.

ACCESSION BC012041 GI:15080597

VERSION BC012041.1

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3845)

Direct Submission

Submitted (30-JUL-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

cDNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalon@bcm.tmc.edu

Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 5 Row: P Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 49893.

FEATURES
Location/Qualifiers

1..3845

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="MGC:6498 IMAGE:2648390"

/tissue_type="Mammary tumor. Metallothionien-TGF alpha

model. 10 month old virgin mouse. Taken by biopsy."

/clone.lib="NCI_GCAP_Mam1"

/lab_host="DH10B"

/note="vector: PCMV-SPORT6"

60..3104

/codon_start=1

/product="Similar to ADP-ribosyltransferase (NAD+; poly

(ADP-ribose) polymerase)"

/protein_id="AAH12041.1"

/db_xref="GI:15080598"

/translation="MAEASRLRYRVEYAKSGRASCCKSCSIPKDSLRMAIMVQSPMF

DGKVPWHYHFCFVKVGHSTIRQPDVEYDGFSELRWDDQKVKTKTAEGAGVAGKQDGS

GGKAETLGDLEAEYAKSNRSMCKGLEIKGQMLSKKRWDPKPKQMGIDRWYHP

TCFVKRDELGERPEYASQSLKGFLLSAEDKALKQLPALKNRGRKRGDEVDGTD

VAKKSKGKDKDSKLEKALKQNELIWNKDELKKACSTNDLKELLIFNQOQVPSG

ESAILDRVADGMGAFALLPKCKESGQLVFKSDAYTCGDTATKCMVKWTONSRKEW

VTPKFEIRISYLKLLKVKQDRIFPPESAPAPLALPLSVTAPTAVNSAPADKKPLS

CDS


```
1464 AGCTTCCAAAGAGCTGCTCTGCGCCACACAGCTTGTCC...TGTGGGGGCG 1510
      ::::: ||| ::::
478 a...LeuGluSerSerLysGly***ThrValThrValLysValLysGlyA 494
      | ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1511 TGAGGTGAAGGAGAGGCTGGTGAAGTGGTGGCCCAAGGGGAAGTCAG 1560
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
494 rSerAlaCysSer***SerPro***ValCysLysAsnThrAlaHisIle 510
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1561 CTGCACCCCTCCAAAGAGAGAGGCTGGTCAAGGAGGAAGGT..... 1604
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
511 Pro****TrpGluLysHisIleGlnCys***LeuLys..... 523
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1605 GTCAACAATCTGAAGAAGAGATGAATTAATCTCAAGGGAGGAGCAGC 1654
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
524 .....HisValLeu..... 526
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1655 CGTTGATCCTGACTCTGGTCTGGAACACTCTGCACACGCTCTGGAGAAAG 1704
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
527 .....Thr***His***Val..... 531
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1705 GTGGGAAGGTGTTACGGCCACACTTGGCTGTGGACATTTGTAAGGG 1754
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
532 CysThrGlyTyrValLeuGlnIleLeuGlnAspGlySerG 548
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1755 ACCAACTCTATTACAACTGCAGCTTCTGGAGGACGACAAAGGAGAGCAG 1804
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
548 uCysTyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyC 565
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1805 GTACTGGATCTTCGGTCTCGGGCGGGTGGGCACA...GTTATCGGCA 1851
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
565 lyGlnLysLeuGluMet...SerLysThrGluAlaIleLysGluPhe 580
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1852 GTACAAACTTGAAGAGATGCCCTCCAAAGAGACGCTGTTGAGCACTTC 1901
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
581 LysArgLeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCy 597
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1902 ATGAAGCTGTATGAAGAGAGAGCTGGGAATGCTGGCACTCG..... 1943
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
597 sLysThrAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValA 614
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1944 ....AAAACCTTCACAAAGATATCCCAAGAAAGTTCTACCCCTCGGAGATTG 1989
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
614 spTyrGlyVal...LysLysAlaProLysArgLysAspIleSerGlu... 628
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1990 ACTATGCCAGGACGAGAGGCTGAAGAGCTGACGGTGAAGCCTGGC 2039
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
629 MetLysSerSerLeuAlaProGlnLeuLeuGluLeuMetLysMetLysPhe 645
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
2040 ACCAAGTCGAAGCTGCCGAAGCCAGTCGAGGAGCTCGTGGGGATGATCTT 2089
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
645 eAsnValGluThrTyrArgAlaAlaMetMetGluPheGlu***AsnMets 662
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
2090 CGAGCTGGAGAGCATGAAGAAGGCTTGGTGGATACGAGATGACCTCC 2139
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
662 erGluMetProLeuGlyLysLysSer*****AsnIleGlu***GlyPhe 678
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
2140 AGAAGATGCCCTTGGGAAGCTGAGCAGAGAGGAGGAGCTCCAGGCGGCTAC 2189
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
679 GluAlaLeuThr*****LeuPheGluGlyHisArg***SerSe 695
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
2190 TCTATCTCTCAGC.....GAGGTCCAGGAGGAGGAGTGC 2221
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
695 rThrGlyLeu***GluLysAla***LeuLeu*****PheS 712
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
2222 TCAAGGAGCAGTGAATCCAGATCTCTCAATCGTCTCTACA 2271
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
712 erLeuLeuSerLeuLeuPheIleLeuIleLeuTyrGlyMetArgMet... 727
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
2272 CTCGTATCCCCATCACTT.....GGAATGAAGAGGCC 2306
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
728 .....IleSerTyrSerLysAlaLysMetLeuG 737
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
```

seq_name: gb_ro:RN094340

seq_documentation_block:

LOCUS RN094340 3128 bp mRNA

19-FEB-1998

DEFINITION Rattus norvegicus poly(ADP-ribose) polymerase mRNA, complete cds.

ACCESSION U94340

VERSION U94340.1 GI:2896791

KEYWORDS

SOURCE

Norway rat.

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 3128)
 Beneke, S., Meyer, R. and Buerkle, A.
 Isolation of cDNA encoding full-length rat (*Rattus norvegicus*) poly
 (ADP-ribose) polymerase
 Biochem. Mol. Biol. Int. 43 (4), 755-761 (1997)
 98046546
 2 (bases 1 to 3128)
 Beneke, S., Meyer, R. and Buerkle, A.
 Direct Submission
 Submitted (18-MAR-1997) Angewandte Tumoriologie, Deutsches
 Krebsforschungszentrum, Im Neuenheimer Feld 242, Heidelberg
 D-69120, Germany
 3 (bases 1 to 3128)
 Beneke, S., Meyer, R. and Buerkle, A.
 Direct Submission
 Submitted (19-FEB-1998) Angewandte Tumoriologie, Deutsches
 Krebsforschungszentrum, Im Neuenheimer Feld 242, Heidelberg
 D-69120, Germany
 Sequence update by submitter
 On Feb 19, 1998 this sequence version replaced gi:2583143.
 Location/Qualifiers
 1..3128
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /cell_type="mononuclear blood cell"
 31..3075
 /codon_start=1
 /product="poly (ADP-ribose) polymerase"
 /protein_id="AAC53544.1"
 /db_xref="GI:2896792"
 /translation="MAEATRLRYVEAKSGRASCCKSEIPKDSIRMAIMVQSPWF
 DGKPVHWHFSCFVGHGSIRODTEVDGPFSELRDDQOKVKTAAAGVAGKQHG
 GGAKETLDPFAEYAKSNSTCKGMEKIEKGMRSLKAMLPDPEKQGLMDIRVHP
 TCFVKNDELGFRPEYASQLKGFSLLSAEDKELKQDPVAKSEKRDQVGDIDE
 VAKSKKDKDESSLEKALKQAQNELVNIKDELKACSTNDLKELLFNQOQVPSG
 ESATLDVADGMFAGLPCQEGSGQLVFKDAYCTGDTAWTKMWKTONPSKEW
 VTKPEFRELKQKLLKQDRLFPPESSAPAPPVSTISAPTAVNSAPADKPLS
 NMKLTLLGSLQNKDEAKMIEKLGKLTGSANKASLCSTKKEVEKMSKKEVKA
 NRVVCEDFLDQYSASAKSLQELLSAHSLSHGAEVKVEGVEVVFPGKSAAPSKSK
 GAVKEGVNKSERKMTLKGGAAYDDSGLESHAVHLEKSGKVFSAITGLVDIVKGT
 NSYKQLLESDESRKTIWFGSRVGTIVGSKLEQMPSEKDAVEHFMKLYEERTGN
 AMHSNKTYPKPYPLEIDYQDEEAVKLVKPGTKSLPKPQVQLGMIFDVESM
 KKALEYEIDLQKPLGKLSRRQIQAYSLSEVQAVSQGSESOILDLNRFYELI
 PHDFGKPKPLNNTDSVQAKVEMNLDDIEVAYSLRGGSDSKDPTDNYEKLK
 TDIKVDRDSEAEVIRKYVKNTHATTHAYDLEVIDIEKIEREGESQRYKPEQLHN
 RLLWHGSRRTNFAGILSGLRIAPPEAVTGYMFGKGIYFADWMSKSNYCHTSQGD
 PIGLILLGEVALGNMYELKHASHISKLPGKHSVKGIGKTAPDPSPASITLDGVEVPLG
 TGPSPGVNDCLLYNEIVYDIAQVNLKLLKLFNFKTSLW"
 BASE COUNT 861 a 758 c 905 g 604 t
 ORIGIN

alignment_scores:
 Quality: 1592.00 Length: 1066
 Ratio: 2.331 Gaps: 36
 Percent Similarity: 64.071 Percent Identity: 36.961

alignment_block:
 US-09-236-995D-2 x RNU94340 ..
 Align seg 1/1 to: RNU94340 from: 1 to: 3128

3 AlaProPolysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAla 19
 |||
 40 GCCACGAGAGGCTTACCGAGTGAGTACCTAAGACGGCGCGCTC 89
 19 rCysLysSerCysArgSerProLysAlaLysAspGlnLeuArgLeu 36
 |||
 90 TTGCAAGAAATGCACGAGAGTATCCCCAGGACTCGCTCGCATGCCA 139

36 ysMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrpAsnHis 52
 |||
 140 TCATGGTGCAGTCACCCATGTTTCGATGGAAAAAGTCCACACTGTTATCAC 189
 53 AlaArgCysIlePheSerLysLysAsnGlnIleLysSerValasp...As 68
 |||
 190 TTCTCTCTGCTTCTGAAGGTGGCCACTCCATCCGGCAGCTTGACACAGA 239
 68 pValGluGlyIleAspAlaLeuArgTrpAspGlnGluLysIleArg 85
 |||
 240 GGTGACGGGTCTCTGAGTCTGGGTGATGACCCAGCAGAGGTCAAGA 289
 85 snTyrVal.....GlySerAlaSerAlaGlyThrSerSerThrAla 98
 |||
 290 AGACCGCAGAGGCTGGAGCGCTGGCAGCAAGCCAGCATGGAGGTGCG 339
 99 AlaProProGluLysCys.....ThrIleGluIleAlaProSe 111
 |||
 340 GGCAGGCGGAGAGACCGCTGGCGACTTTGGCGCAGAGTATGCCAAGTC 389
 111 rAlaArgThrSerCysArgCysSerGluLysIleThrLysGlySerV 128
 |||
 390 CAACAGGAGCACATGCAAGGCTGCATGGAGAGATAGAGAGGTCAGA 439
 128 alArgLeuSerAlaLys...LeuGluSerGluGlyProLys...GlyIle 142
 |||
 440 TCGGCTGTCCAAGAGATGCTGGATCCAGAGAGCCGCAACTGGGTATG 489
 143ProTyrHisAlaAsnCysPheGlu..... 152
 |||
 490 ATTGACCGCTGGTACCATCCCACTTGTTCAGAAACCGGGACGAGCT 539
 153ValSerProSerAlaThrValGluLysPheSerGlyTrpAspThrL 168
 |||
 540 GGGCTTCGGCTCAGTACACAGGCTGTCAGCCAGCTCAAGGGCTTTTGGCTCC 589
 168 euSerAspGluAspLysArgThrMet.....LeuAspLeuValLys 181
 |||
 590 TCTCTGCAGAGACAAAGAGCTCTGAAGAGAGAGCTCCCGGCGAGTCAAG 639
 182 LysAspValGlyAsnAsnGluGlnAsnLysGlySerLysArgLysLys 198
 |||
 640 AGTGAA.....GGAAAGAGAGAAATGTGA 662
 198 rGluAsnAspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrS 215
 |||
 663 CGAGCTGGAT.....GGAATAGATGAA..... 684
 215 erGluGlyThrValArgAsnLysGlyGlnLeuValAspProArgGlySer 231
 |||
 685GTAGCCAAAAGAAATCTAAA.....AAAGGAAAG 714
 232 AsnThrSerSerAlaAspIleGlnLeuLysLysGluGlnSerAspTh 248
 |||
 715 GACAAGGAGAGTAGTAACCTGGAAAAAGCGCTCAAGGCTCAGAACGAGCT 764
 248 rLeuTrpLysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGlu 265
 |||
 765 GGTCTGGAATATCAAGATCAGCTGAAGAAAGCTCTCCACCAACGACC 814
 265 euArgAspMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArg 281
 |||
 815 TGAAGGAGTGTCTCATCTTCAACAGCAGCAGTACCATCTGGAGAGTCC 864
 282 HisLeuLeuAspArgCysAlaAspGlyMetLeuPheGlyAlaLeuGly 298
 |||
 865 GCATCTGACCGAGTGTCTGATGCTGGCGTGGCGCTTGGGGCCCTCTTCC 914
 298 oCysProValCysAlaAsnGlyMetTyrTyrTyrAsnGlyGlnTyrGlnC 315
 |||
 915 CTCAGAGGAGTGTTCAGGACAGCTGGTGTAAAGAGCAGCGCTTATTACT 964
 315 ysSerGlyAsnValSerGluTrpSerLysCysThrTyrSerAlaThrGlu 331

1823 GTAACAAGCTTGACGATGCCGTCCAAAGAGGATGCTGTGGAACACTTC 1872
581 LysArgLeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCy 597
1873 ATGAAGTTATACGAAGAAAGACTGGAANTGCCCTGGCACTCG 1914
597 sLysThrAsnPheArgLysGlnProGlyArgPheTrpProLeuAspVala 614
1915AAAAACTTCACAAAGTATCCCAAGAAAGTTCTACCCCTCGGAGATTG 1960
614 sPyrGly.....ValLysLysAlaProLysArgLysAsp 625
1961 ACTATGGCCGACGACGAAGAGCAGTAAAGAGCTGGCGTGAAGCCTGGC 2010
626 IleSerGluMetLysSerSerLeuAlaProGlnLeuLeuGluLeuMetLy 642
2011ACTAAGTCGAAGCTGCCGAAGCGGTGCAGGAGCTCGTCGG 2051
642 sMetLeuPheAsnValGluThrTyrArgAlaAlaMetMetGluPheGlu* 659
2052 GATGATCTTCGACGTGGAGCATGAAGAGCCCTTGGTGAGTACGAGA 2101
659 ***AsnMetSerGluMetProLeuGlyLysLeuSer*****AsnIleGlu 675
2102 TTGACCTTCAGAAAGATGCCCTTGGGGAACACTGAGCAGAGGAGCATCCAG 2151
676 ***GlyPheGluAlaLeuThr*****LeuPheGluGlyHisAr 692
2152 GCGGCTACTCTATCTCAGT.....GAGGTCCAGCA 2183
692 g***SerSerThrGlyLeu***GluLysAla***LeuLeu***** 709
2184 GGCAGTGTCCCAAGGAGCAGTGAATCTCAGATCCTGGATCTCTCCAATC 2233
709 *****PheSerLeuLeuSerLeuLeuPheIleLeuIleLeuTyrGlyMet 725
2234 GCTTACACACCTCATCCCCCATGACTTT.....GGAATG 2268
726 ArgMet.....IleSerTyrSerLysAlaLys 734
2269 AAGAAGCCCTTCTGAAACAACAGACAGTGTGCAGGCCAAGGTGA 2318
734 sMetLeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleVal.... 749
2319 GATGCTGGACACCTCTCGACATCAGGTGGCTATAGCTTCTTCAGGG 2368
750 ..GlyPheAspSerAspSerAspSerLeuAspAspLysTyrMetLys 765
2369 GTGGGTCTGATGACAGCAGTAAGGACCCCATCGATGTCAACTATGAGAAG 2418
766 LeuHisCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLe 782
2419 CTCAAAACTGACATTAAAGGTGGTTGACAGAGATTCTGAAGAGCTGAGGT 2468
782 uIleGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrps 799
2469 CATCAGGAAGTATGTGAAGAACACTCAGCGCACCCACCAATGCCCTATG 2518
799 erLeuGluLeuGluValPheSerLeuAspArgAspGlyGluLeuAsn 815
2519 ACCTGGAAGTGATAGACATCTTTAAGATAGAGCGAGAGGAGAGCCAA 2568
816 LysTyrSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHi 832
2569 CGGTACAAAGCCCTTCAGG...CAGCTTCAACACCGGAGACTGCTGTGGCA 2615
832 sGlySerArgLeuThrAsnPheValGlyIleLeuSerGlnClyLeuArgI 849
2616 CGGTCCAGGACCAACCAACTTCGAGGAGCATCTGTACAGGGTCTGGGA 2665
849 leAlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeu 865
2666 TAGCCCACTGGAAGCACCTGTGACAGGCTACATGTTTGGGAAAGGAATC 2715

965 GTACTGGGATGCTCACTGCTGACCAAGTGTGCTCAAGACACAGAAAT 1014
332 ProValArgValLysLysTrpGlnIleProHisGlyThrLysAsnAs 348
1015 CCAGCCGA.....AAGGAATGGGTAACTCCA.....AAGGAATT 1049
348 pTyrLeuMetLysTrpPheLysSerGlnLysValLysLysProGluArgV 365
1050 CCAGAAATATCTACCTCAAGAAATTAAGATCAAAAGACAGGACCGGC 1099
365 alleuProPrometSer.....ProGluLysSerGlySer 376
1100 TGTTCCTCCAGAGACAGCGCCCGACACCAAGCAGCCTCCGCTCG 1149
377 LysAlaThrGlnArgThrSerLeuLeuSerSerLysGlyLeuAspLys... 392
1150 ATCAGCTCCGACCCACAGCTGTGAAGTCTCTCGCACAGCAGACAAAGC 1199
393LeuArgPheSerValValGlyGlnSerLysG 403
1200 CTTGCTTAACATGAAGATCTGACTCTGTGAAGCTCTCCCAAGCAAGG 1249
403 luAlaAlaAsnGluTrpIleGluLysLeu.....LysLeuAlaGlyAla 417
1250 ACGAAGCAAGGCATGATTGCAAACTCGAGGCAAGCTGACAGGATCT 1299
418 AsnPheTyrAlaArgValValLysAspIleAspCysLeuIleAlaCysGl 434
1300 GCCAACAAGGCTCTCTG.....TGTATCAGCATGAAGAA 1334
434 yGluLeuAspAsnGluAsnAlaGluAlaArgLysAlaArgLeuLysI 451
1335 GGAGGTGGAGAAGATGATGAAGATGGAGAGTGAAGAGTGAAGAGCAGCCAA 1384
451 leProIleValArgGluGlyTyrIleGlyGluCysValLysArgThrLys 467
1385 TTCAGTGTGTGAGGACTTCTCCAGGATGTCTGCTGCTCAGCTAAG 1434
468CysCysHisLeuIleCysIleAsnTrpAsnAl 478
1435 AGCTCCCAAGAGCTGCTCTCAGCCCATAGTGTGCC...TCGTGGGGGC 1481
478 a...LeuGluSerSerLysGly***ThrValThrValLysValLysGlyA 494
1482 TGAGGTGAAGTAGAGCTGTGAAGTGGTGGTCCCCCAAGGGAGGATCAG 1531
494 rgSerAlaCysSer***SerPro***ValCysLysAsnThrAlaHisIle 510
1532 CTGCACCTCCCAAGAGAGAGGAGGTGTGTCAAGGAGGAAGGT..... 1575
511 Pro*****TrpGluLysHisIleGlnCys***LeuLys..... 523
1576 GTCAACAAGTCTGAAGAGAGATGAATAACTACTGTAAGGGAGGAGCAGC 1625
524HisValLeu..... 526
1626 CTTGATCTGACTGTGCTGGAACACTCTGCACATGCTCTGGAGAAAG 1675
527Thr***His***Val..... 531
1676 GCGGAAGGTGTTCAAGTCCAGCGCTGGTGGTGGACATTGTGAAGGG 1725
532 CysThrGlyTyrTyrValLeuGlnIleIleLeuGlnAspAspGlySerGl 548
1726 ACTAAGTCTTATTAACAACTGCAACTCTGGAGTCCGACAGGAGGAGCAG 1775
548 ucystyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyG 565
1776 GTATTGGATCTTCGGGTCTCGGGCCGAGTGGGCACA...GTATATCGCA 1822
565 lyGlnLysLeuGluMet....SerLysThrGluAlaIleLysGluPhe 580

866 TyrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspAr 882
 |||||
 2716 TACTTTGCTGATATGCTGTCACAAAGTCCGAACACTACTCCACACGCTCA 2765
 |||||
 882 gAsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspM 899
 |||||
 2766 GGGAGACCCGATGCTGCTAACTGTTGGGAGAAAGTTGCCCTTGGAACA 2815
 |||||
 899 eTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLys 915
 |||||
 2816 TGTACGAACACTCAAGCATGCTTCTCACATCAGCAAGTTACCAAGGGCAAG 2865
 |||||
 916 HisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVa 932
 |||||
 2866 CACAGTGTCAAGAGTTTGGGCAAAACCGCCCTGACCCCTTCGGCCACGAT 2915
 |||||
 932 lLysTrpArgAspValValValProCysGlyLysProValProSerS 949
 |||||
 2916 CACC...CTGGATGCTGTAGAGGTTCGCGTGGGAACAGGGATTCCGTCG 2962
 |||||
 949 eRIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThr 965
 |||||
 2963 GTGTTAATGACACCTGCTGCTGTATACAGGATACATTTGCTACGACATT 3012
 |||||
 966 SerGlnValLysMetGlnPheLeuLeuLysValArgPheHisLys 981
 |||||
 3013 GCTCAGGTGAATCTGAAGTACCTACTGAAACTGAAGTTCAATTCAAG 3060
 |||||

seq_name: gb_pat:114359

seq_documentation_block: 3747 bp DNA PAT 26-SEP-1995
 LOCUS 114359
 DEFINITION Sequence 5 from patent US 5449605.
 ACCESSION 114359
 VERSION 114359.1 GI:996850
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 3747)
 Smulson,M.E., Lyn,D. and Cherney,B.
 TITLE Method of detecting a predisposition to cancer by detecting a
 deletion polymorphism in the gene for human poly (ADP-ribose)
 polymerase
 JOURNAL Patent: US 5449605-A 5 12-SEP-1995;
 FEATURES
 Location/Qualifiers
 1..3747
 /organism="unknown"
 BASE COUNT 1036 a 844 c 1060 g 807 t
 ORIGIN

alignment_scores:
 Quality: 1587.00 Length: 1060
 Ratio: 2.290 Gaps: 36
 Percent Similarity: 65.377 Percent Identity: 36.226

alignment_block:
 US-09-236-995D-2 x 114359 ..
 Align seg 1/1 to: 114359 from: 1 to: 3747

6 LysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLysSe 22
 |||||
 209 AAGCTCTATCGAGTCTGAGTACGCAAGAGCGCGCGCTCTTGCAAGAA 258
 |||||
 22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetValG 39
 |||||
 259 ATGCAGCAGACAGACATCCCAAGAGACTCGCTCGGATGGCCATCATGTGTC 308
 |||||
 39 lnaLaserGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
 |||||

309 AGTCGCCCATGTTTATGGGAAAAGTCCACACACTGGTACCACCTTCTCTCTGC 358
 |||||
 56 IlePheSerLysLysAsnGlnIleLysSerValAsp...AspValGluGl 71
 |||||
 359 TTTCTGGAAGGTGGGCCACTCTCATCCGCCACCTTGACGTTGAGGTGGATGG 408
 |||||
 71 yIleAspAlaLeuArgTrpAspAspGlnGluLysIleArgAsnTyrVal. 87
 |||||
 409 GTTCTCTGACCTTCGGTGGATGACCAGCAGAAAGTCAAGAGACAGCGG 458
 |||||
 88GlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
 |||||
 459 AAGCTGGAGGAGTGCACAGCAAGGCCAGGATGGAATTGGTAGCAGGCA 508
 |||||
 102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
 |||||
 509 GAGAAAGACTCTGGGTGACTTTTCAGCAGCAGATATGCCAAGTCCCAACAGAAG 558
 |||||
 114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLeu 131
 |||||
 559 TAGGTCCAAAGGCTGTATGGAAGAAGTAAAGAGGCCAGGTGCCCTGT 608
 |||||
 131 eAlaLysLeu...GluSerGluGlyProLys...GlyIle.....Pro 143
 |||||
 609 CCAAGAAGATGTTGGACCCGAGAACACACACTAGGCATGATTGACCGC 658
 |||||
 144 TrpTyrHisAlaAsnCysPhePheGlu.....ValSe 154
 |||||
 659 TGGTACCATCCAGGCTGCTTCTCAAGAACAGGAGGAGCTGGGTTTCCG 708
 |||||
 154 rProSerAlaThrValGluLysPheSerGlyTrpAspThrLeuSerAspG 171
 |||||
 709 GCCCGAGTACAGTGGCGATCAGCTCAAGGGCTTCAGCCTCTCTGTGTACAG 758
 |||||
 171 luAspLysArgThrMetLeuAspLeuValLysLysAspValGlyAsnAsn 187
 |||||
 759 AGGATAAA.....GAAGCCCTGAAGAAGCAGCTCCAGGAGTC 796
 |||||
 188 GluGlnAsnLysGlySerLysArgLysLysSerGluAsnAspIleAspSe 204
 |||||
 797 AAG.....AGTGAAGGAAGAAAGAAAGCGGATGAGGTGGATGG 834
 |||||
 204 TrpLysSerAlaArgLeuAspGluSerThrSerGluGlyThrValArg 221
 |||||
 835 A.....GTGGATGAAGTGGCGAAGAAGAAATCTAAAAAAG 869
 |||||
 221 snLysGlyGlnLeuValAspProArgGlySerAsnThrSerSerAlaAsp 237
 |||||
 870 AAAAAGACAAG.....GATAGTAAG 889
 |||||
 238 IleGlnLeuLysLysGluLysSerAspThrLeuTrpLysLeuLysAs 254
 |||||
 890 CTTGAAAAGCCCTAAAGGCTCAGAACGACCTGATCTGGAACATCAAGGA 939
 |||||
 254 pGlyLeuLysThrHisValSerAlaAlaGluLeuArgAspMetLeuGlu 271
 |||||
 940 CGAGCTAAAGAAAGTGTGTTCACTAATGACCTCAAGGAGCTACTCATCT 989
 |||||
 271 laAsnGlyGlnAspThrSerGlyProGluArgHisLeuLeuAspArgCys 287
 |||||
 990 TCAACAAGACCAAGTCCCTTCTGGGAGTGGCGGATCTTGGACCCAGTA 1039
 |||||
 288 AlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAlaAs 304
 |||||
 1040 GCTGATGGCATGGTGTTCGGTGGCTCTCTTCCCTGCGAGGAATGCTCGGG 1089
 |||||
 304 nGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerGlyAsnValSerG 321
 |||||
 1090 TCAGCTGGTCTTCAAGAGCATGCTATTACTGCACCTGGGAGCTCACTG 1139
 |||||
 321 luTrpSerLysCysThrTyrSerAlaThrGluProValArgValLysLys 337
 |||||
 1140 CTGGACCAAGTGTATGGTCAAGACAGACAGACCCCAACCGG.....AAG 1183
 |||||

[illegible]

```
2894 TCACAGAGTGCCAACTACTACATACGCTCTCAGGAGACCAATAGGCTT 2943
888 uMetLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysA 905
2944 AATCCTGTTGGGAGAGTGGCCCTTGGAAACATGTATGAATGAAGCAGC 2993
905 laThrSerMetAspLysProProArgGlyLysHisSerThrLysGlyLeu 921
2994 CTTACATATACAGCAGTATGCCAAGGCAAGCAGTGTCAAAGTTTG 3043
922 GlyLysThrValProLeuGluSerCluPheValLysTrpArgAspVa 938
3044 GCCAAACTACCCCTGATCTTCAGCTAACATTAGT...CTGGATGGTG 3090
938 lValValProCysGlyLysProValProSerSerIleArgSerSerCluL 955
3091 AGACGCTTCTCTGGGCGGATTCATCTGCTGCTGCTGCTGCTGCTGCT 3140
955 euMetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 971
3141 TACTATATAAGCAGTACATTTCTATGATATGCTCAGGTAAATCTGAAG 3190
972 PheLeuLeuLysValArgPheHisHisLys 981
3191 TATCTGCTGAACCTGAATCAATTTTAAG 3220
seq_name: gb_ro:MMADPRP
seq_documentation_block:
LOCUS MMADPRP 3172 bp mRNA ROD 12-SEP-1993
DEFINITION Mouse mRNA for poly (ADP-ribose) polymerase (BC 2.4.2.30).
ACCESSION X14206
VERSION X14206.1 GI:49893
KEYWORDS DNA-binding protein; NAD(+) ADP-ribosyltransferase; zinc finger protein.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 3172)
Huppi.K.
Direct Submission
Submitted (26-JAN-1989) Huppi K., NCI/NIH, Lab of Genetics, Building 37, Room 2B-21, Bethesda, MD 20892
2 (bases 1 to 3172)
Evers.R., Hammer.A. and Cornelissen.A.W.C.A.
Unusual C-terminal domain of the largest subunit of RNA polymerase II of Chithidia fasciculata
Nucleic Acids Res. 17, 3387-3401 (1989)
Location/Qualifiers
1..3172
/organism="Mus musculus"
/strain="BxSB"
/db_xref="taxon:10090"
/tissue_type="spleen"
/clone="pADP3-2.1 and pADP-3C"
85..3126
/notes="poly (ADP-ribose) polymerase (AA 1 - 1013)"
/codon_start=1
/protein_id="CAA32421.1"
/db_xref="GI:49894"
/db_xref="MGI:87933"
/db_xref="SWISS-PROT:P11103"
/translation="NAEASRLRYQVAKSGRACKKSESIPKDSLRMAIMVQSPMF
DGKVPVHYHFCFMKVGSIKHPDVEGFSRLRWDDQKVKKTAEGAGVAGKQDGS
GGKAETLGDPAEAKNSRMCKGLEIKGQMRISKMWDPKPKQLGMIDRWYHP
TCFVKRDELGRPYTSASQLKFSLLSAEDKALKQLPAIKNEGRKRGDEVDGTE
VAKKSRDELTGKLEKALKAQNELIWNIDELKACSTNDKLELIFNOQOVPSGE
SAILDRVADGMAFGALLPCKECSGOLVFKDAYCTGDTAVTKCMVKTQNPCKEHW
TPKEFEITSYLLKLVKQDRIFPESSAPITVHVPLSVTSAPTVAVNSAPADPKPSN
MKILTLGLSONKDEKAVIEKLGKLTGSANKASLCISIKKEVERNNKMEVKAN
IRVSEDFLDQVSASTKSLQDLLSAHLSLSPWGAEPVAVPRGKSAAPSKKSG
```

```
CFKEGVNKSSEKRMKLTTLKGAADVDPDSGLEHSAHVLEKGGKVSATLGLVDIVKGTN
SYTKLQLEDDKESRYWIFRSWGRCLCTVIGSNKLRMPKSEAEVORPMKLYEKTGNA
WHSKNETKYPKPKFYPLEIDYGGDEAVKLTVPKGTKSLPKPQELVGMIFDVDSMK
KALVEYEIDLOKMPGLKSRROIQAAVSTLSEVQPVQSGSSEQLDLSNRFYTLIP
HDFGMMKPPLLANNADSVQAKVEMLNLDIEVAYSLRGGSDSDSKDDIDVNEKLLT
DIKVDPRDSEAEVIRKYKYNTHATTHNAYDLVIEDIFIEREGESQRYKPFQRLHNR
RLMHGSRITTFAGLISQGLRIAPPEAVPTGYMGFGIYFADWVSKSANYCHTSGQDP
IGLIMIGEVALGNMYELHSHASKLPRKHSVGLGKTTDPSPASITLEGVEVPLGT
GIPSGVNDTALLYNEYIVYDIAQVNLKYLKLNKFNKTSLW"
88..678
misc_feature /note="zinc finger domains"
2743..2823
misc_feature /note="DNA-binding domain A"
2898..2952
misc_feature /note="DNA-binding domain B"
BASE COUNT 869 a 779 c 925 g 599 t
ORIGIN
alignment_scores:
Quality: 1584.50 Length: 1081
Ratio: 2.340 Gaps: 37
Percent similarity: 62.627 Percent identity: 36.078
alignment_block:
US-09-236-995D-2 x MMADPRP
Align seg 1/1 to: MMADPRP from: 1 to: 3172
3 AlaProProLysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlase 19
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
94 GCCTCGGAGAGGCTTTTTCGAGTCAGTACGCAAGAGCGGCGCGCTC 143
19 rCysLysSerCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyL 36
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
144 TTGCAAGAAATGCAGCAGAGATATTCCCAAGCACTCCCTCCGATGGCCA 193
36 YsMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrpAsnHis 52
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
194 TCATGTGTGAGTCCCATTTTCGATGGGAAAGTCCACACTGGTATACC 243
53 AlaArgCysIlePheSerLysLysAsnGlnIleLysSerValAsp...As 68
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
244 TTCTCTCTGCTTCTGGAAGTGGCGCAGTCCATCCGCGCACCTGATGTTGA 293
68 pValGluGlyIleAspAlaLeuArgTrpAspGlnGluLysIleArgA 85
: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
294 GTGTGATGGCTTCTCTGAGCTCGCGTGGGATGATCAGCAGAAAGTCAAGA 343
85 snTyrVal.....GlySerAlaSerAlaGlyThrSerSerThrAla 98
: : : : : : : : : : : : : : : : : : : : : : : : : :
344 AGACGCGCGAGCTGGAGCGTGGCAGCAAGCGCAGGATGGAAGTGGG 393
99 AlaProProGluLysCys.....ThrIleGluIleAlaProSe 111
: : : : : : : : : : : : : : : : : : : : : : : : : :
394 GGCAAGCGCGAGAGACATTTGGTGACTTTTGCAGCAGATGATGCCAAGTC 443
111 rAlaArgThrSerCysArgCysSerGluLysIleThrLysGlyServ 128
: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
444 CAACAGGAGCATGTGCAAGGCGCTGCGAGAGAGATAGAGAAGGCGCAGA 493
128 alArgLeuSerAlaLysLeu...GluSerGluGlyProLys...GlyIle 142
: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
494 TGCCTCTGCTCCAGAAAGATGGTGGATCCAGAGAACCCACAGCTGGGTATG 543
143 .....ProTrpTyrHisAlaAsnCysPhePheGlu..... 152
: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
544 ATTGACCGCTGGTACCATTCCAACTGCTGCTTTGTCAAGAAAGCGGACGAGCT 593
153 ....ValSerProSerAlaThrValGluLysPheSerGlyTrpAspThrL 168
: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
594 GGGCTTCCGCGCGCAGTACAGTCCAGCTCAGCTCAGCTTACAGGCTTTAGCCTCC 643
```

[illegible]

```

2240 TGCTCAAGCGCAGGTGAATCCAGATCTCCAGATCTAGATCTCTCCAATCGCTTC 2248
711 PheSerLeuLeuSerLeuLeuPheIleLeuIleLeuTyrGlyMetArgMe 727
      ::::||||: ||| |||||:
2290 TACACTCTGATCCCCCATGACTTT .....GGATGAGAA 2324
727 t.....lleSerTyrSerLysAlaLysMetL 736
2325 GCCCCCACTCCTGAACAACGACAGCGTGCAGGCGCAAGTGGAGATGC 2374
736 euGluAlaLeuGlnAspIleGluIleLeuSerLysIleVal.....Gly 750
      ||:: ||| |||||:||||: ||| |||||:
2375 TAGACAACCTCTCGACATCGAGGTGGCCTATAGTCTCTTCAGGGGTGC 2424
751 PheAspSerAspSerAspGluSerLeuAspAspLysTyrMetLysLeuHi 767
      |||:||||: ||| :::: |||||: |||||: |||||:
2425 TCTGAGCAGCAGCAGGATCCATCGAGCTCAACTACGAGAACTCAA 2474
767 sCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeuLeG 784
      |||||: :::: |||||:||||: ||| |||||:
2475 AACTGCATTTAAGGTGTGACAGAGATTCTGAAGAGCGCGAGGTGCATCA 2524
784 luGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSerLeu 800
      ::::||||: |||||:||||: |||||: |||||:
2525 GGAAGTACGTGAAGAACACTCATGCTACCACGCACAAACGCTATGACCTG 2574
801 GluLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTy 817
      ::::||||: |||||:||||: |||||: |||||:
2575 GAAGTGATCGATATCTTCAAGATAGAGCGGAGGGGAGAGCGCAGCGCTA 2624
817 rSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlyS 834
      :::: |||||: |||||: |||||: |||||: |||||:
2625 CAAGCCCTTCAGG...CAGCTTCACAACCGAGGCTGCTGGCAGGGCT 2671
834 erArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAla 850
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2672 CCAGGACCACCAACTTTGCTGGCATCTGTGCAGGGTCTGCGGATAGCC 2721
851 ProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPh 867
2722 CCACCTGAAGCCCGCTGACAGCTACATGTTGGGAAGGGATCTACTT 2771
867 eaLaspLeuValSerLysSerAlaGlnTyrCysTyrValAspArgAsnA 884
2772 TGCCGACATGTTGTCCAAAAGTGCAAACTACTGCGCACACATCTCAGGGAG 2821
884 snProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMetTyr 900
2822 ACCCGATTGGCTTAATAATGCTGGAGAGGTGGCCCTTGGAAATATGAT 2871
901 GluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysHisSe 917
2872 GAACTCAAGCATGCTTCATCATATCAGCAAGTTACCCAGSGCGAAGCACAG 2921
917 rThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheValLysT 934
2922 TGTCAAAGGTTTGGGAAAAACCCACCCCTTGACCCTTCGGCCAGCATCAC 2970
934 rPArgAspAspValValProCysGlyLysProValProSerSerIle 950
2971 ..CTGGAGGGTAGAGGTTCCACTGGGAACAGGATCCCATCTGGGTGC 3018
951 ArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSerGl 967
3019 AACGACACTGCCCTGCCTGTATTAATGAGTACATGTTCTACGACATGCTCA 3068
967 nValLysMetGlnPheLeuLysValArgPheHisHisLys 981
3069 GGTGAATCTCAAAATACCTGCTGAAACCTCAAGTTCAATTTTAA 3111
seq_name: gb_pat:AX058340
seq_documentation_block:

```

seq_name: qb pat:AX058340

seq_documentation_block:

| | | | | | |
|------------|--|-------------|-------------|-----|-------------|
| LOCUS | AX058340 | 3045 bp | DNA | PAT | 17-JAN-2001 |
| DEFINITION | Sequence 24 | from Patent | WO00771179. | | |
| ACCESSION | AX058340 | | | | |
| VERSION | AX058340.1 | GI:12310800 | | | |
| KEYWORDS | human.. | | | | |
| SOURCE | Homo sapiens | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 3045) | | | | |
| AUTHORS | Christenson,E., Demaggio,A.J., Goldman,P.S. and Mcelligott,D.L. | | | | |
| TITLE | Human poly(adp-ribose) polymerase 2 materials and methods | | | | |
| JOURNAL | Patent: WO00771179-A 24 21-DEC-2000; | | | | |

| FEATURES | source | CDS | ORIGIN |
|-----------------------|---|-----|--------|
| Location (Qualifiers) | | | |
| 1. 3045 | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| 1. 3045 | | | |
| | /note="unnamed protein product" | | |
| | /codon_start=1 | | |
| | /protein_id="CAC22488.1" | | |
| | /db_xref="GI:12310801" | | |
| | /translation="MAESDKLVREYAKSGRACKCKSESIPKDSLRWALIMVQSPF | | |
| | DGKVPWYHPSCFWYKVGHSIRHPDVEDGFSELRWDDQKVKKTARAGGVYTGKQDGI | | |
| | GSRAEKTLDGFAEAYKSNRSTCGCKMEKIERQVRLSKRMVDPEKPKQLGMDIRWYHP | | |
| | GCVFKNREIEGPRPEYSASLQKGSILLATIEDKALKKQLQPGVSEKRGKDEVDGVE | | |
| | VAKKIKKEDKDKSLKEKALKAQNDLIWNIKELKKVCTNDLKELLIFNKQVPQSGE | | |
| | SATLDRVADMGFGALLPCBEGSQGLVFKSDAYICGTDVMTAPAAVNSASADKPLSN | | |
| | TPKEFEISYLLKKVKVDRIPEPTASVAATPPPTASAPAAVNSASADKPLSN | | |
| | MKTLILGKSLNDEVKAMIEKLGKLTCTANKASLCISTKKEVEKMNKKMEEVKRAN | | |
| | IRVVEDFDQDSASTKSIOBLEFLAHLILSPWGAEVRAEPEVAVPAPKSGAALSKRSK | | |
| | GKVEKGNSEKKRMKLTILKGGAAVDPDGSLHSAHLEKGGKVFATSLGLVDIVKGT | | |
| | NSYKLIQLLEDENRNIWIPRSGNRVTGVIISGNLQKMPKSKEDATBEQMKLVEETGN | | |
| | AMHSKNTFYKPKFLEIDYGODEAEAVKRLTPNGCTKSLKPKVPQDILKMTFDVESH | | |
| | KKAMVFEIDLOKPLGKLSKRGTOAAYSILSEVQASVQSGSDSOILDLNSRFYTLI | | |
| | PHDFGMKPKPLLNADSVQAKFEMLDNLLDIEWAYSLRGSGDDSKDPIDVNYEKLL | | |
| | TDILRWGRDSEEAETIRKYVKNTHATHSYADLEVIDFIEREGECGRKPKPKQLHN | | |
| | RRLIWHGSTRNFAGLISQGLRTATPEAPVTGMFGKGIYFADMWKSKSNANYHTSQGD | | |
| | ILRLIGLVEALGNMYELKHSAISLPRGKHSVRLGKGTTPDPSANISLDGDVDPVLPG | | |
| | TGISSGVDSILNYEIVYDIAQVNLKYLKKLKNFKTSLM" | | |
| | 863 a 710 c 851 g 621 t | | |

```

alignment_scores:
  Quality: 1584.00      Length: 1060
  Ratio: 2.289          Gaps: 36
  Percent Similarity: 65.283      Percent Identity: 36.226

alignment_block:
  US-09-236-995D-2 x AX058340      ..

Align seg 1/1 to: AX058340 from: 1 to: 3045

```

```

6   LysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLysSe 22
    ||| :||| :||||| :||||| :||||| :||||| :||| :||| :|||
19  AAGCTCATCGAGTCGAGTAGCCCAAGAGCGGC GGCCCTTTGTGAAGA 68
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
22  rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetValG 39
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
69  ATGCAGCGAGGAGATCCCCAAGGACTCGCTCCGGATGGCCATCATGGTG 118
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
39  lnAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
119 AGTCGCCCATGTTTGATGGAAAGATCCCACACTGGTACCATTCTCCTGC 168
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
56  IlePheSerLysLysAsnGlnIleLysSerValasp... AspValGlucI 71
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 TTCTGGAAAGTGGGCACATCCATCCGCGCACCTCGAGCTTCAGGTTGATGG 218
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
71 ylleAspAlaLeuArTrpAspAspGlnGluLysIleArgAsnTyrrVal.. 87
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

[illegible]

Mon Mar 11 09:58:37 2002

```
617 al...LysLeuAlaProLysArgLysAspLeuSerGlu...MetLysSer 631
      :::::||||| :::::
1940 AGGATGAAGAGCGCAGTGAAGAAGCTCAGATAAATCTGGCAGCAAGTCC 1989
      :::::||||| :::::
632 SerLeuAlaProGlnLeuLeuGluLeuMetLysMetLeuPheAsnValG1 648
      :::::||||| :::::
1990 AAGCTCCCAAGCCAGCTTCAGAGCTCATCAAGATGATCTTTGATGGGA 2039
      :::::||||| :::::
648 uThrTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetP 665
      :::::||||| :::::
2040 AAGTATGAAGAAAGCCATGCTGGAGATGATGATGATGATGATGATGATG 2089
      :::::||||| :::::
665 toLeuGlyLysLeuSer*****AsnLeGlu***GlyPheGluAlaLeu 681
      :::::||||| :::::
2090 CCTTGGGGAAGCTGAGCAAAAGGAGATCCAGCGCATATCTCCATCCTC 2139
      :::::||||| :::::
682 Thr*****LeuPheGluGlyHisArg***SerSerThrGlyLe 698
      :::::||||| :::::
2140 AGT.....GAGGTCCAGCAGCGGTGCTCAGGGCAG 2171
      :::::||||| :::::
698 u**GluLysAla***LeuLeu*****PheSerLeuLeuS 715
      :::::||||| :::::
2172 CAGCAGCTCTCAGATCCTGGATCTCTCAATCGCTTTTACACCTGATCC 2221
      :::::||||| :::::
715 erLeuLeuPheLeuLeuLeuLeuTyrGlyMetArgMet..... 727
      :::::||||| :::::
2222 CCCAGCAGCTTT.....GGGATGAAGAGCGCTCCGCTCCTG 2256
      :::::||||| :::::
728 .....IleSerTyrSerLysAlaLysMetLeuGluAlaLeuG1 740
      :::::||||| :::::
2257 AACAAATGACAGACAGTGTCCAGGCCAAGTGGAAATGCTTGACAACTGCT 2306
      :::::||||| :::::
740 nAspIleGluIleAlaSerLysIleVal.....GlyPheAspSerAsps 755
      :::::||||| :::::
2307 GCACATCAGGTGGCTCAGCTGCTCAGGGAGGGTCTCATGATAGCA 2356
      :::::||||| :::::
755 erAspGluSerLeuAspAspLysTyrMetLysLeuHisCysAspIleThr 771
      :::::||||| :::::
2357 GCAAGATGCCATCGATGCTCAACTATGAGAAGCTCAAAACTGACATTAA 2406
      :::::||||| :::::
772 ProLeuAlaHisAspSerGluAspTyrLysLeuIleGluGlnTyrLeu 788
      :::::||||| :::::
2407 GTGGTTGACAGAGATCTGGAAGAGCCGAGATCATCAGAGATGATTAA 2456
      :::::||||| :::::
788 uAsnThrHisAlaProThrHisLysAspTrpSerLeuGluLeuGluGlu 805
      :::::||||| :::::
2457 GAACACTCATGCAACACACACAGTGCCTGATGACTTGAAGTCAATCGATA 2506
      :::::||||| :::::
805 aPheSerLeuAspArgAspGlyGluLeuAsnLysTyrSerArgTyrLys 821
      :::::||||| :::::
2507 TCTTTAAGATAGAGCGTGAAGCGAATGCCAGCGTTACAAGCCCTTTAA 2556
      :::::||||| :::::
822 AsnAsnLeuHisAsnLysMetLeuLeuThrPheHisGlySerArgLeuThr 838
      :::::||||| :::::
2557 ...CAGCTTCATAACCGAAGATGCTGTGGCAGCGGTCCAGAGCACCAC 2603
      :::::||||| :::::
838 nPheValGlyIleLeuSerGlnGlyLeuArgIleAlaProProGluAla 855
      :::::||||| :::::
2604 CTTTCTGGGATCCTGTCCTCAGGCTTCGGATAGATCCCGCTGAGCGC 2653
      :::::||||| :::::
855 roValThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuVal 871
      :::::||||| :::::
2654 CCGTGACAGCTACATGTTTGGTAAAGGATCTATTTCGCTGACATGTC 2703
      :::::||||| :::::
872 SerLysSerAlaGlnTyrCysTyrValAspArgAsnProValGlyLe 888
      :::::||||| :::::
2704 TCCAGAGTGCCCACTTACCTAGCTCAGGGAGACCCCAATAGGCTT 2753
      :::::||||| :::::
888 uMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLys 905
      :::::||||| :::::
2754 AATCCTGTTGGGAGAGTGGCTTGGAAACATGATGATGAAGCAGC 2803
      :::::||||| :::::
905 laThrSerMetAspLysProProArgGlyLysHisSerThrLysGlyLeu 921
```

```
||||| :::::||||| :::::||||| :::::||||| :::::
2804 CTTCAATATCAGCAGTTACCAAGGGCAACACAGTGTCAAGGTTG 2853
      :::::||||| :::::
922 GlyLysThrValProLeuGluSerGluPheValLysTyrArgAspAspVa 938
      :::::||||| :::::
2854 GGCAAACTACCCCTGATCTTTCAGCTAACATAGT...CTGGATGCTGT 2900
      :::::||||| :::::
938 lValValProCysGlyLysProValProSerSerIleArgSerSerGlu 955
      :::::||||| :::::
2901 AGACGTTCTCTTGGGACCGGATTCATCTGCTGATAGACACCTCTC 2950
      :::::||||| :::::
955 euMetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 971
      :::::||||| :::::
2951 TACTATATACAGTACATGCTATGATATTCCTCAGGTAATCTGAAG 3000
      :::::||||| :::::
972 PheLeuLeuLysValArgPheHisHisLys 981
      :::::||||| :::::
3001 TATCTGCTGAAGCTGAATTCATTTTAAG 3030
      :::::||||| :::::
seq_name: gb_pat:AX062277
seq_documentation_block: 3045 bp DNA PAT 24-JAN-2001
LOCUS AX062277
DEFINITION Sequence 136 from Patent WO0100849.
ACCESSION AX062277
VERSION AX062277.1 GI:12540178
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3045)
AUTHORS Christenson, E., Demaggio, A.J., Goldman, P.S. and Mcelligott, D.L.
TITLE Tankyrase2 materials and methods
JOURNAL Patent: WO 0100849-A 136 04-JAN-2001;
ICOS CORPORATION (US)
FEATURES
source
1..3045
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3045
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25068.1"
/db_xref="GI:12540179"
/translation="MAESSDKLYRVEYAKSGRASCCKKCSSEIPKDSLRLMAITWQSPMF
DGKVPWHYHFSCFWKVGSHIRHPDVEDVDGFSSELWDDQKVKRTAEAGVGTGKQDGI
GSKAEKTLGDFAAEYAKSNRSTCKGMEKTEKGQVRLSKMVDPEKPOLGMDIRWYHP
GCFVKNREELGFRPEYSASQLKGFSLIATDEKALKKQLPGVKSEGRKGDVGVDE
VAKKSKKEKDKSKLEKALAONDILWNIKDELKVCSTNDLKELLIFNKKQVPSGE
SAILDRVADGWVFGALLPCESCSGLVFKSDAYCTGDTATWTKMVKTOIPNKEWV
TPKEPREISYLKKLVKKODRIFFPETSASVAANTPSTASAPAAVSSASADKPLN
MKILTLGLSRNKDEVKAMIEKLGKLTGTANKASCLISTKKEVKMKMKKEEYKSK
TRVSEDFLDQVSATSKIQLEFLAHLSPWGAEPVEPVEVAPRGKSGAALSKSK
GOVKEEGLNKSEKRMKLTGKGAADVDPDSGLEHSAHVLKGGKVEDAEQPMKLYEEKTN
NSYKLOLLEDDKENRYVIGSWRGVTVIGSNLEQMPKEDAEIQPMKLYEEKTN
AWHSNFTYKPKFPLEIDYQDEEAVKLTVPNPKTKSLPKPKVDQDLIKMIFVESM
KKAMVEYEDIQMPLGKLSRQIAATSIILSEVOQAVSQSSSSQILDLNSRYPTLI
PHDFGMKPPPLNNADSOAKVEMDNLDDLEAVSLRGGSDSSDDIPDVTNTEKLR
TDIKVDRDSEAEIIRKYVNTHATSHAYDLEVIDIFKIEREGECORPKPKOLHN
RELLWHGSRITNFAGILSQGLRAPPAPVTGYMFGKGIYFADMYVSKSANYHTSQGD
PIGLLILGEVALGNMYELKHASHISRLPKGHSVKGKLTTPDPSPANISLGDVDVPLG
TGISSGVDTSLYNEYIVYDIAQVNLKYLKLNFKNTSLW"
```

BASE COUNT 863 a 710 c 851 g 621 t
ORIGIN

alignment_scores:

Quality: 1584.00 Length: 1060
Ratio: 2.289 Gaps: 36
Percent similarity: 65.283 Percent identity: 36.226

alignment_block:

US-09-236-995D-2 x AX062277

Align seg 1/1 to: AX062277 from: 1 to: 3045

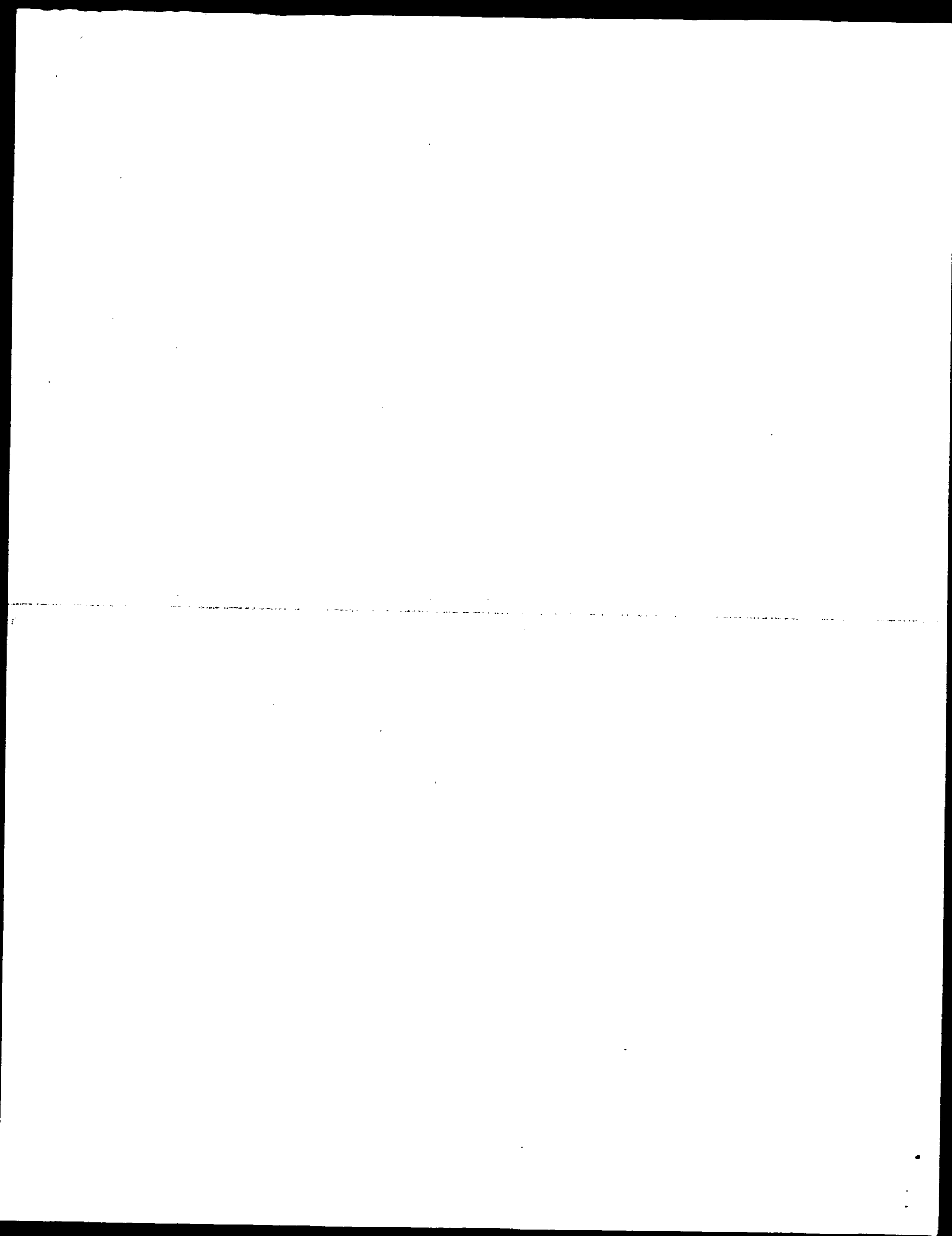
6 LysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLys 22
19 AAGTCTATCGAGTCGAGTCAGCAAGAGCGGCGCCCTCTTGCAGAA 68
22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetValG 39
69 ATGCAGCAGAGCATCCCAAGGACTCGCTCCGATGGCCCATCATGGTGC 118
39 InAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
119 AGTCGCCCATGTTGATGTAAGAAAGTCCACACATGGTACCACTTCTCTGC 168
56 IlePheSerLysLysAsnGlnIleLysSerValAsp...AspValGluG1 71
169 TTCGGAAGTGGGCCACTCCATCCCGCACCCCTGACGTTGAGGTGGATGG 218
71 yIleAspAlaLeuArgTrpAspGlnGluLysIleArgAsnTyrVal. 87
219 GTTCTCTGAGCTTCGGTGGATGACACAGAGAAAGTCAAGAACACACGG 268
88GlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
269 AAGCTGGAGGAGTGCACAGCAAGGCCAGGATGGAATTGGTAGCAAGGCA 318
102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
319 GAGAAGACTCTGGGTGACTTTGTCAGCAGAGATATGCCAAGTCCACAGAA 368
114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLys 131
369 TACGTGCAAGGGTGTATGAGAGATAGAAAAGGCCAGGTGCGCCTGT 418
131 erAlaLysLeu...GluSerGluGlyProLys...GlyIle.....Pro 143
419 CCAAGAAGATGGTGGACCCGAGAGAGCCACAGCTAGGCATGATGACCGC 468
144 TrpTyrHisAlaAsnCysPhePheGlu.....ValSe 154
469 TGGTACCATCCAGGCTGCTTTGTCAAGAACAAGGAGGAGTGGGTTTCG 518
154 rProSerAlaThrValGluLysPheSerGlyTrpAspThrLeuSerAsp 171
519 GCCCGAGTACGTCGAGTACGTCAGGCTTCCAGGCTTCAGCCTCTCTGAC 568
171 LuAspLysArgThrMetLeuAspLeuValLysLysAspValGlyAsnAsn 187
569 AGGATAAA.....GAAGCCCTGAAGAAGCAGCTCCAGGATC 606
188 GluGlnAsnLysGlySerLysArgLysSerGluAsnAspIleAspSe 204
607 AAG.....AGTGAAGGAAGAGAAAAGCGCATGAGGTGGATGG 644
204 rTyrLysSerAlaArgLeuAspGluSerThrSerGluGlyThrValArgA 221
645 A.....GTGGATGAAGTGGCGAAGAAGAAATCTAAAAAAG 679
221 snLysGlyGlnLeuValAspProArgGlySerAsnThrSerSerAlaAsp 237
680 AAAAAGACAAG.....GATAGTAAG 699
238 IleGlnLeuLysLeuLysGluLysSerAspThrLeuTrpLysLeuLysAs 254
700 CTTGAAAAGGCCCTAAAGGCTCAGAAGCACCCTGATCTGGAACATCAAGGA 749
254 pGlyLeuLysThrHisValSerAlaAlaGluLeuArgAspMetLeuGluA 271
750 CGAGCTAAAGAAAGTGTCTCAACTAATGACCTGAAGGAGCTACTCATCT 799
271 laAsnGlyGluAspThrSerGlyProGluArgHisLeuLeuAspArgCys 287

800 TCAACAAGCAGCAAGTGCCTCTGGGAGTGGCGATCTTGGACCGAGTA 849
288 AlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAlaAs 304
850 GCTGATGGCATGGTGTTCGTCCTCTTCCCTCGGAGGAATGTCGGG 899
304 nGlyMetTyrTyrTyrAsnGlyGlnTyrGlnCysSerGlyAsnValSerG 321
900 TCAGCTGGTCTTCAAGAGCGATGCTTACTTGCACATGGGAGCTCACTG 949
321 LuTrpSerLysCysThrTyrSerAlaThrGluProValArgValLysLys 337
950 CTGGGACCAAGTGTATGGTCAAGACACAGACACCAACCGG.....AAG 993
338 LysTrpGlnIleProHisGlyThrLysAsnAspTyrLeuMetLysTrpPh 354
994 GAGTGGGTAAACCCA.....AAGGAATTCGAGAAATCTCTTACT 1034
354 eLysSerGlnLysValLysLysProGluArgValLeuProMetSer. 370
1035 CAAGAAATGAAGTTAAAGAGCAGACCGTATATTCCCCCCAGAAACCA 1084
371ProGluLysSerGlySerLysAla 378
1085 GGGCCPCCGTGGCGGCCACGCTCCGCCCTCCACAGCCTCGGCTCTGCT 1134
379 ThrGlnArgThrSerLeuLeuSerLysGlyLeuAspLysLeuArgPh 395
1135 GCTGTGAATCTCTCTCTTCCAGCAGATAGCCATATCCACATGAAGAT 1184
395 eSerValValGly.....GlnSerLysGluAlaAlaAsnGluTrpI 409
1185 CCTGACTCTCGGAAGCTCTCCCGGAACAAGATGAAGTGAAGGCCATGA 1234
409 leGluLysLeu.....LysLeuAlaGlyAlaAsnPheTyrAlaArgVal 423
1235 TTGAGAACTCGGGGGGAAGTTGACGGGACGCCCAACAAGGCTCCCTCG 1284
424 ValLysAspIleAspCysLeuIleAlaCysGlyGluLeuAspAsnGluAs 440
1285TGCAATCAGCACCAAAAAGGAGGTGGAAAAGATGA 1319
440 nAlaGluValArgLysAlaArgLeuLysIleProIleValArgGluG 457
1320 TAAGAAGATGGAGGAAGTAAGGAAGCAACATCCGAGTTGTCTGAGG 1369
457 lyTyrIleGlyGluCysValLysArgThrLys..... 467
1370 ACTTCTCCAGGACGTCCTCCGCTCCACCAAGAGCCTTCAGGAGTTGTT 1419
468 CysCysHisLeuIleCysIleAsnTrpAsnAlaLeuGluSerSerLysG1 484
1420 TTAGCCACATCTGTCC...CCTTGGGGGCGAGAGGTGAAGCAGAGCC 1466
484 y***ThrValThrValLysValLysGlyArgSerAlaCysSer***SerP 501
1467 TGTGAAGTT...GTGGCCCCAAGAGGGAAGTCAGGGGCTGGGCTCTCCA 1513
501 ro***ValCysLysAsnThrAlaHisIlePro****Trp..... 514
1514 AA.....AAAAGCAAGGCCAGGTCAAGGAGGAAGGTATCAACAAA 1554
515 ...GluLysHisIleGlnCys***LeuLys..... 523
1555 TCTGAAAAGAGAAATGAATTAACCTTAAGAGGAGGACGCTGTGATCC 1604
524HisValLeu..... 526
1605 TGATTCTGGACTGAACACTCTCGCATGTCTCTGGAGAAAGGTGGGAAG 1654
527Thr***His***Val.....CysThrGly 534

1655 TCTTCAGTCCCAACCGTTGGGCTGTGCATCCTTAAGAAGAACCAACTCC 1704
535 TyrTyrValLeuGlnIleGluAlaAspGlySerGluCysTyrVa 551
||||| ||||| ||||| ||||| ||||| :||| :||| :||| :|||
1705 TACTACAAGCTGCAGCTTGAGGAGCACAGAAACAGATTGGAT 1754
551 lPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLysL 568
:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
1755 ATTCAAGTCTCTGGGCCGCTGGGGTAGG...GTGATCGGTAGCAACAAC 1801
568 euGlucUmet...SerLysThrGluAlaIleLysGluPheLysArgLeu 583
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
1802 TGGAACAGATGCCGTCFCAAAGAGGATGCCATTCAGCATTCATGAATAA 1851
584 PheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAs 600
:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
1852 TATGAAGAAAACCGGGAACGCTTGGCACTCC.....AAAA 1889
600 nPheArgLysGlnProGlyArgPheTrpProLeuAspValAspTyrGlyV 617
||||| ||| ||| ::|||::|||::|||::|||::|||::|||::|||:
1890 TTTCACGAAGTATCCCCAAAAGTTTTACCCTCGAGATGACTATGGCC 1939
617 al...LysLysAlaProLysArgLysAspLysSerGlu...MetLysSer 631
::::::::::|||::|||::|||::|||::|||::|||::|||::|||:
1940 AGGATGAAGAGCGCATGAAGAAGTCAAGTAAATCCCTGSCACCAAGTCC 1989
632 SerLeuAlaProGlnLeuGluLeuMetLysMetLeuPheAsnValGI 648
::::|||::|||::|||::|||::|||::|||::|||::|||::|||:
1990 AAGCTCCCACAGCCAGTTTCAGGACCTCATCAAGATATCTTGTATGGGA 2039
648 uThrTyrArgAlaLaMetMetGluPheGlu***AsnMetSerGluMetP 665
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
2040 AAGTATGAAGAAAGCCATGGTGGAGTATGAGATCGACCTTCAGAGATCC 2089
665 roLeuGlyLysLysSer*****AsnIleGlu**GlyPheGluAlaLeu 681
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
2090 CCTTGGGAAAGCTGAGCAAAAGCCAGATCCAGCCGCATACCTCATCTC 2139
682 Thr*****LeuPheGluGlyHisArg***SerSerThrGlyLe 698
::|||::|||::|||::|||::|||::|||::|||::|||::|||:
2140 AGT.....GAGGTCCAGCAGCGGTCTCTCAGGGCAG 2171
698 u**GluLysAla***LeuLeu*****PheSerLeuLeus 715
|||::|||::|||::|||::|||::|||::|||::|||::|||:
2172 CAGCAGCTCTCAGATCTCGGATCTCTCAANTCGCTTTACACCTTGATCC 2221
715 erLeuLeuPheIleLeuLeuTyrGlyMetArgMet..... 727
|||::|||::|||::|||::|||::|||::|||::|||::|||:
2222 CCCAGCACTT.....GGGATGAAGAAGCGCTCCGCTCTC 2256
728lleSerTyrSerLysAlaLysMetLeuGluAlaLeucl 740
2257 ACAATGCGACAGATGTGCAGGCCAAGGTGGAAATGCTTGACAACCTGCT 2306
740 naSPileGluIleAlaSerLysIleVal.....GlyPheAspserAsps 755
|||||::|||::|||::|||::|||::|||::|||::|||::|||:
2307 GGACATCGAGTGGCTCAGCTGCTCAGGGGAGGGTCTGATGATAGCA 2356
755 eAspGluSerLeuAspLysTyrMetLysLeuHisCysAspIlethr 771
|||::|||::|||::|||::|||::|||::|||::|||::|||:
2357 GCAAGGATCCCATGATGTCAACTATGAGAAGCTCAAAACTGACATTAAG 2406
772 ProLeuAlaHisAspSerGluAspTyrLysLeuIleGluInTyrLeuLe 788
|||::|||::|||::|||::|||::|||::|||::|||::|||:
2407 GTGGTGTGACAGAGATCTGAAGAAGCGGAGATCATCAGGAAGTATGTTAA 2456
788 uaSnThrHisAlaProThrHisLysAspTrpSerLeuLeuGluGluV 805
|||||::|||::|||::|||::|||::|||::|||::|||::|||:
2457 GAACACTCATCCACCACACACAGCTGCGTATGACTTGGAGTCATCGATA 2506
805 alPheSerLeuAspArgSpGlycyluLeuAsnLysTyrSerArgTyrlys 821
:|||::|||::|||::|||::|||::|||::|||::|||::|||:
2507 TCCTTTAAGATAGAGCGTTGAAGCGAATGCGCAGGTTTACAGCCCTTAA 2556

us-09-236-995d-2.rge

Mon Mar 11 09:58:37 2002




```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAspPheValGlyIle 17
|||||
2549 AACAGAGAGCTATTATGCGACGCTTCAAGGTTCAGCAATTTTGGGGAAT 2598
17 eleuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||||
2599 TCTTAGTCAAGGGCTAAGATTGCACTCCCTGAGGACCTTTCTGCGCT 2648
34 yrmelPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerLysSerAla 50
|||||
2649 ATATGTCGCGAAGGCTCTACTTTCAGATCTAGTAAGCAAGAGCGCA 2698
51 GlnTyrrCysTyrrValAspArgAsnAsnProValGlyLeuMetLeuLeu 67
|||||
2699 CAATACCTGTTATGTGATAGGAATAATCTGTAGCTTTGATGCTTCTTC 2748
67 rgluValAlaLeuGlyAspMetTyrrGluLeuLysLysAlaThrSerMet 84
|||||
2749 TGAGGTCCTTAGAGACATGATGAACTTAAGCAAAACCGTCATG 2798
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
2799 ACAAACTCCAAAGAGGCAATTCGACCAAGGATTAGCGCAAAACCGTG 2848
101 ProLeuGlnSerGluPheValLysTrpArgAspValValProGly 117
|||||
2849 CCACGTGAGTCAGAGTTGTGAGAGTGAGGATGATGTGAGTTCCTG 2898
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrrAsn 134
|||||
2899 CGGCAAGCGGTCATCATCAATAGAGCTCGAACTCATGTACAATG 2948
134 luTyrrIleValTyrrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
2949 AGTACATGCTCTACCAACATCCCGAGTGGAAGATCGAGTCTGTGTAAG 2998
151 ValArgPheHisHisLysArg 157
|||||
2999 GTGCTTTCATCACAAAGG 3019
seq_name: gb_p1:AF093627
seq_documentation_block:
LOCUS AF093627 3285 bp mRNA PLN 29-NOV-1998
DEFINITION Zea mays poly(ADP)-ribose polymerase (PARP1) mRNA, complete cds.
ACCESSION AF093627
VERSION AF093627.1 GI:3928870
KEYWORDS
SOURCE
ORGANISM
Zea mays.
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 3285)
REFERENCE
AUTHORS Mahajan,P.B. and Zhu,Z.
TITLE Purification and cDNA cloning of maize Poly(ADP)-ribose polymerase
JOURNAL Plant Physiol. 118 (3), 895-905 (1998)
MEDLINE 99026291
REFERENCE
AUTHORS Mahajan,P.B. and Zhu,Z.
TITLE Direct Submision
JOURNAL Submitted (22-SEP-1998) Crop Protection, Pioneer HiBred
International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131-1004,
USA
FEATURES
Source
1..3285 Location/Qualifiers
/organism="Zea mays"
/db_xref="taxon:4577"
1..3285
/gene="PARP1"
100..3042
/gene="PARP1"
/EC_number="2.4.2.30"

```

```

/function="add ADP-ribose"
/codon_start=1
/product="poly(ADP)-ribose polymerase"
/protein_id="AAC79704.1"
/db_xref="GI:3928871"
/translation="MAAPPKMAKAEYAKSGRASCSRSPIARDQRLKMKVQASQFD
GFEMPMNARCIFSKNOIKSVDDVEGIDALRMDDEKIRINVGASAGTSYAAPPE
KCTEIAPSARTSCRSCEKITKSGVLESAKLESEEGKGIIPMYHANCFFEVSAAYE
KFSQMPTLSDERKTM.DLYKKDVGNNKNGKSKKSENDISYSARLDESTSGT
VRNGQLVDPGCSSTSSADTOLAKRESDPLMKLKKRKYHVAALRDLKLEANGDT
SGPRHLDRADMDLFGALGPCVCANMGT.YINGOTQCGANVSENSKCTYSATEPR
VKRWQIPHGTRKNDYLMKWKFSOKVKKPERVLPMSPESKGSATORTSLSSGLDK
LRFVSGQSKAANEMWIEKLKLAGANVYARVYKIDICLAGELDENAVRARRLK
LPIVREGIGECVKKNMKLPEDLYKLENALESSGSTVYKVGSRVHVSGLDPTA
HILPDDCKSTVMAATIMSDIALGVNSYVLOI.IEDDQSECEYFRKMGVSEKIGQK
LEBKSITBAIKERFLRLKLETKGNSWMECKTNRRQGRFPLDNDVGVKAPKRD
ISEKSSIAPLDLELMKLEFVETRYRAMMEFELINSMELPKLSKENIKGFEALTE
IONLKIDPADQALAVRESLI.VAASNREFTLIPSTIPIH.IDEDDLMTKAMBALDOI
ELASIKVGFSDSDSDSLDDKMKLHCDITPLADSEDDYKLIBOYLNTNAPTRKDWSL
ELEEVFSLDRDGEINKYSRYKNNLHNMKLLMHGSRFLNFGYILSOGIRIAPPAVYTG
YMRKGLYFADIVSKSAOYCVDDRNPNVGLMLSEVALGMYELKATKATSMDKPRKH
STKGIGKTVPILESEFVKKRVDVVPCKPSPSSIRSELWYNEIYVNTSQVKMQLL
KVRPHHKR"
BASE COUNT 1003 a 617 c 828 g 837 t
ORIGIN

```

```

alignment_scores:
Quality: 828.00 Length: 157
Ratio: 5.274 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-236-995d-4 x AF093627

Align seq 1/1 to: AF093627 from: 1 to: 3285

```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAspPheValGlyIle 17
|||||
2569 AACAGAGAGCTATTATGCGACGCTTCAAGGTTCAGCAATTTTGGGGAAT 2618
17 eleuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||||
2619 TCTTAGTCAAGGGCTAAGATTGCACTCCCTGAGGACCTGTTAGTGGCT 2668
34 yrmelPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerLysSerAla 50
|||||
2669 ATATGTCGCGAAGGCTCTACTTTCAGATCTAGTAAGCAAGAGCGCA 2718
51 GlnTyrrCysTyrrValAspArgAsnAsnProValGlyLeuMetLeuLeu 67
|||||
2719 CAATACCTGTTATGTGATAGGAATAATCTGTAGTGTGATGCTTCTTC 2768
67 rgluValAlaLeuGlyAspMetTyrrGluLeuLysLysAlaThrSerMet 84
|||||
2769 TGAGGTCCTTAGAGACATGATGAACTTAAGAAAGCCACGCTCATGG 2818
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
2819 ACAAACTCCAAAGAGGCAATTCGACCAAGGATTAGCGCAAAACCGTG 2868
101 ProLeuGlnSerGluPheValLysTrpArgAspValValProGly 117
|||||
2869 CCACGTGAGTCAGAGTTGTGAGAGTGAGGATGATGTGATGCTCCCTG 2918
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrrAsn 134
|||||
2919 CGGCAAGCGGTCATCATCAATAGAGCTCGAACTCATGTACAATG 2968
134 luTyrrIleValTyrrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
2969 AGTACATGCTCTACCAACATCCAGGTGAAGATCGAGTCTGTGTAAG 3018

```

151 ValArgPheHisLysArg 157
 |||
 3019 GTGGCTTCCATCAAGAGG 3039

seq_name: gb_pl:ATH131705

seq_documentation_block: 3187 bp mRNA 18-DEC-1998
 LOCUS Arabidopsis thaliana mRNA for poly(ADP-ribose) polymerase.
 DEFINITION Arabidopsis thaliana mRNA for poly(ADP-ribose) polymerase.
 ACCESSION AJ131705
 VERSION AJ131705.1 GI:4038490
 KEYWORDS NAD(+) ADP-ribosyltransferase; parp-1 gene; poly(ADP-ribose) polymerase.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 3187)
 Kazmaier, M.
 Direct Submission
 Submitted (17-DEC-1998) Kazmaier M., Departement D'Ecophysiologie Vegetale Et De Microbiologie, Commissariat A L'Energie Atomique, CEA Cadarache, Lab. De Radiobiol. Vegetale, Bat. 185, 13108 St. Paul-les-Durance, FRANCE
 2 (bases 1 to 3187)
 Doucet-chabeaud, G. and Kazmaier, M.
 Unpublished
 Location/Qualifiers
 1..3187
 /organism="Arabidopsis thaliana"
 /cultivar="landsberg erecta"
 /db_xref="taxon:3702"
 11..2963
 /gene="parp-1"
 11..2963
 /gene="parp-1"
 /standard_name="NAD(+) ADP-ribosyltransferase"
 /EC_number="2.4.2.30"
 /function="ADP-Ribose polymer synthesis"
 /note="mol. weight: 111 kDa"
 /codon_start=1
 /evidence=experimental
 /product="poly(ADP-ribose) polymerase"
 /protein_id="CAA10482.1"
 /db_xref="GI:4038491"
 /translation="MASPHKPPWRAEYAKSSRSCKTCKSVINKENFRLGKLVOSTHFD
 GIMPMWNASCLLKTQIKSDVDVEGIESLRWEDQOKIRKYVESGAGSNTSTSGIS
 TSSANNAKLEYGLEVQTSRAGCRKCEKILGEVRIFSKPEGPNGKGLMWHAKCF
 LEMSSELESLSGWSRIPDSQDALLPLVKALPAAKTAETAEARQTNRSAGTKRND
 SVNKEKSLAKSSFDMSGALQPCSKEMEATKELWDLKDDLKRYVTSAEIRML
 EVNQSTRGSELDLRKACDGMFPLALCPMCSGHLSPGGLYRCHYIYSEWSKCSH
 STLPDRIKGKWKIPDETENQFLKWNKSOKSVKPKRILRPVLSETSGOGSKDAD
 SSRERLADLVKSTAGNTPROPHKPIEAGAEHANVKGTSCVLVCGSLDIRDAE
 MKRARKVAIRDYLVDFCKQKLPDFKIKEDTSESLVTIVKGRSAVHEASGL
 QERCHILGDNSTLTMSDSTGINSYIIQIQEDKGSDCYFVRKWRGNEKI
 GKNVEMSKSDAVHEFKRLFEKTGNTWSEQKTNFQKPGKFLPLDIDYGNKQV
 AKKPFQTSNLSPLLEMLMFLDVTYRSAMFEINMSEMPGLGKSKHNKQGF
 ALTIQRLLTESDQPTMESLLVDASNRFETWIPSTPHIIRDEDDFKSKVKMLAL
 QDIASRIKVGFDVSTESLDKYLKLCDDI SPLPHDSERYRLIEKVLNTHAPTE
 WSLREVEFALEREGERDKYAPHEKLGKMLNHWGSRUNTFVGLNQGLRIAPPEAP
 ATGVFKSGIYFDLVSKAQCYCTCKNPVGLMLLSEVALGEIHELTRAKYMDKPPR
 GKHSKGLGKKNVPDSEFAKWRGDTVPCGKPVSSKVKASELMYNEYIVYDTAQVQLQ
 FLKVRFEKRR"
 14..2959
 /gene="parp-1"
 /EC_number="2.4.2.30"
 /product="poly(ADP-ribose) polymerase"

gene

CDS

source

mat_peptide

BASE COUNT 1044 a 577 c 743 g 823 t

alignment_scores:

Quality: 697.00 Length: 157
 Ratio: 4.709 Gaps: 0
 Percent Similarity: 94.268 Percent Identity: 82.166

alignment_block:

US-09-236-995D-4 x ATH131705 ..

Align seg 1/1 to: ATH131705 from: 1 to: 3187

1 AsnLysMetLeuLeuThrPheHisGlySerArgLeuThrAsnPheValGlyI 17
 |||
 2489 AATAAGATGCTCTCTATGGCATGGTTCTCGATTAAACGAATTTTGTGGAT 2538

17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
 |||
 2539 ATTGAACCAAGACATGAGNATTCACCTCCAGAACTCTGCTACTGGTT 2588

34 yMetPheGlyLysGlyLeuThrPheAlaAspLeuValSerIysSerAla 50
 |||
 2589 ACATGTTTGGAAAGGGATATATCTTTGCTGACCTTGTCTAGTAAAGTCT 2638

51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeu 67
 |||
 2639 CAGTACTGCTACACTGTGAAGAAAAATCCGGTGGTCTAATGCTTCTGAG 2688

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMet 84
 |||
 2689 TGAAGTTGATTGGGAGAAATACATGAGCTACAAAAGCTAAGTATATGG 2738

84 sPLysProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
 |||
 2739 ATAAACCTCCGAGAGGAAACACATCGACCAAGGGCTCGCAAGAAAGTG 2788

101 ProLeuGluSerGluPheValLysTyrPargAspValValProCy 117
 |||
 2789 COTCAAGATTCCGAGTTTCCCAAGTGGAGAGTGATGTGACTGTTCCTCG 2838

117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsn 134
 |||
 2839 TGGAAACCTGTTTCATCAAGGTCAAGGCTCTGAGCTTATGTACATG 2888

134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLys 150
 |||
 2889 AGTATATGCTCTACGATACAGCCAGGTGAAGTTGCAGTTCTTGTGAAA 2938

151 ValArgPheHisLysArg 157

2939 GTAAGTTTAAAGCACAGAGA 2959

seq_name: gb_ro:RNPARP2

seq_documentation_block:

LOCUS RNPARP2 1939 bp DNA ROD 20-MAY-1992
 DEFINITION R.norvegicus mRNA for poly(ADP-ribose) polymerase (clone PRATC).

ACCESSION X65497

VERSION X65497.1 GI:56849

KEYWORDS NAD(+) ADP-ribosyltransferase.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (23-MAR-1992) F. Potvin, Molecular Endocrinology, CHUL

[illegible]

REFERENCE 1 (bases 1 to 3845)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 5 Row: p Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 49893.
Location/Qualifiers
1. .3845
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:6498 IMAGE:2648390"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI-CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
60. .3104
/codon_start=1
/product="Similar to ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)"
/protein_id="AAH12041.1"
/db_xref="GI:15080598"
/translation="NAEASRLRYVEYAKSGRACKCKSESIPKDSLRMAIMVQSPMF DQVPHVHFCFVKVGHSDIROPDVEVDFSELWDDQKVKVKTAEAGVAGKGDGHS GKAETLGDFAEYAKSNRSMCKGCKLEKQMLSKKWKVDPEKPOLGMDIRWHP TCFVKKRDELGFPEYSASQLKGFSLISAEDREALKQLPAIKNEGKRGKDEVDGTE VAKKSKGDKDSSKLEKALKAKNELIWNIDELKACSTNDLLELILFNQOQVPSG ESAILDRVADGAFGALLPCKECQGLVFSKDAYCTGDTWTMTKMTQNPSSKEW VTPKEPREISYLKLVKKQDRIFPPSSAPAPLALPLSVTSAPTAVNSAPADKPLS NKILTLGLKLSQNKDEAKAV IEKLGKLTGSANKASLCISTKKEVKKSKMEEVKAA NVRVCEDELQDVASASTLSQELLSNHSLSMWGAEVKAPPEVAPKCKSAAPSKKS GAVKEGVNKSERKMKLILKGGAADPDGSLGHSARHVLKGGKVSATLGLVDIVKGT NSYIKQLLEDDEKSYWIFRSWGRVGTIGVSKLEQMSPKDEDAVEHFMKLYEERTGN AWHSKNFTYKPFYPLEIDYGDDEAVKLTVPKTKSLPKPQVIELVGMFTFVESH KVALVEYIDKQMPGLKLSRQIOAAYSILSEVOAVSQGSESIQILDSNRYFTLI PHDGMKPEPLLNADSVQAKVEMDLNLDIEVAYSLIRGSDSDSKDPIDVNYEKLK TDLKVYDRSEAEVIRKYVKNTHATTHAYDLEVIDLPIKIEREGESORYKPFROLHN RLLWHSRTITFAGLISQGLRTAPEVDTGMFGKGLYFADVMKSKSNYCHTSQGD PTGLILLGALVGNLHSHISKLPGKSHVKGKLTPTDPPSASITLLEGVEVPLG TGIPSGVNDTCLLYNIYIVDYTAQVNLKLLKLFNFKTSLW"

BASE COUNT 1042 a 930 c 1080 g 793 t
ORIGIN

alignment_scores:
Quality: 518.50 Length: 156
Ratio: 3.812 Gaps: 1
Percent Similarity: 87.179 Percent Identity: 62.179

alignment_block:
US-09-236-995d-4 x BC012041

Align seg 1/1 to: BC012041 from: 1 to: 3845

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17
|||||
2625 AACCGAGGCTCTGTGGCAGCGCTCCAGGACCACTTGTGGCAT 2674
|||||
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||
2675 CCTGTGCGAGGCTCTGCGGATAGCCCACTGAAGCGCTGTGACAGGCT 2724
|||||
34 yMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
2725 ACATGTTGGGAAAGGATCTACTTTGCGGACATGGTGTCCAAAAGTGCA 2774
|||||
51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSe 67
|||||
2775 AACTACTGCCACACATCTCAGGAGACCGGATGGCTTAATACTGCTGGG 2824
|||||
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||
2825 AGAGGTGTCCTTGGAAACATGTATGAACATCAAGCATGCTTCACATATCA 2874
|||||
84 sPlsProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
2875 GCAAGTTACCAAGGGCAAGCACAGTGTCAAAGGTTGGGAAAAACACC 2924
|||||
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||||
2925 CTGACCTCTGGCCACGATCACC...CTGGAGGGGTAGAGGTTCCACT 2971
|||||
117 sGlyLysProValProSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||
2972 GGAACAGGAGATCCCATCTGGTGTCAACGACACCTGCCTGCTGATAATG 3021
|||||
134 LuTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
|||||
3022 AGTACATTTGTTACGACATGCTCAGGTGAATCTCAATACCTGCTGAA 3071
|||||
151 ValArgPheHisLysLys 156
|||||
3072 CTCAGTTCAATTTTAAG 3089

seq_name: gb_pr:HUMADPPO

seq_documentation_block:
LOCUS HUMADPPO 1771 bp mRNA PRI 30-OCT-1994
DEFINITION Human placental poly(ADP-ribose) polymerase mRNA, partial cds.
ACCESSION M17081
VERSION M17081.1 GI:178151
KEYWORDS Human placenta cDNA to mRNA, clone lambda-PAP803.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1320; 1560 to 1771)
AUTHORS Suzuki,H., Uchida,K., Shima,H., Sato,T., Okamoto,T., Kimura,T. and Miwa,M.
TITLE Molecular cloning of cDNA for human poly(ADP-ribose) polymerase and expression of its gene during HL-60 cell differentiation [published erratum appears in Biochem Biophys Res Commun 1987 Nov 13:148(3):1549]
JOURNAL Biochem. Biophys. Res. Commun. 146 (2), 403-409 (1987)
MEDLINE 87298455
REFERENCE 2 (bases 1321 to 1559)
AUTHORS Suzuki,H., Uchida,K., Shima,H., Sato,T., Okamoto,T., Kimura,T. and Miwa,M.
TITLE Errata
JOURNAL Biochem. Biophys. Res. Commun. 148, 1549-1550 (1987)
FEATURES
1. .1771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="lq41-q42"

```

gene      1. .1725
          /gene="ppol"
CDS       <1. .1725
          /gene="ppol"
          /note="poly(ADP-ribose) polymerase"
          /codon_start=1
          /db_xref="GDB:G00-119-508"
          /protein_id="AAA51599.1"
          /db_xref="GI:178152"
          /translation="KKMEVKEANIRVVSEDFLDYASASTKSLQELFLAHLSPWCAE
          VKAPVEVAPRGSKGAALSKSGQVKEGINKSEKMKLTLLKGAADDPDPSGLSHS
          AHWLEQKGFVSATLGLVDIVKGTNSYKQLLEDKKNRYWIFRSGRWGTVIGSNK
          LEQMSKEDANEHFMKLYBEKTNAMSKNFTYKPKFYPLEIDYGDDEAVKLTVN
          PGTKSKPKFQVQDLIKMIFDVESMKRAMVEYIDLOKPLGLSKRQIOAAYSILSEV
          QOAVSGSSDQILDLNRFYTLIPHDGFMKKPPLNNADSVQAKVEMLDNLLDIEVA
          YSLLRGSDSSDDPDIDVNYEKLTKDKVYDROSEAEIIRKYVKNTHATTHNAYDLE
          VIDFKLERGEQORYKPKFOLNRRLLHWSRTTNFAGLSQGLRIAPPEAPVTGYM
          FGIKGFADWVSKSANYCHPTSQDPTGLILLGEVALGNMYELKHASHISLKPGRHSV
          KGLGKTTDPFSPANISLDGVDVPLGTGISSGNDTSLLYNEYIIVDYAQNLYLLK
          FNFKTSLW"

BASE COUNT      501 a 409 c 472 g 389 t
ORIGIN

alignment_scores:
  Quality: 517.50      Length: 156
  Ratio: 3.777        Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:
US-09-236-995D-4 x HUMADPPO ..
Align seg 1/1 to: HUMADPPO from: 1 to: 1771
1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
1246 AACCGAAGATTGCTGTGGCAGCGGTCCAGGACCAACATTTTGTCTGGAT 1295

17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
1296 CTTGTCCAGGCTCTCGGATAGCCCGCCCTGAAGCCGCCGTGACAGGCT 1345

34 yMetPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerLysSerAla 50
1346 ACATGTTTGGTAAGGATCTATTTCGCTGACATGGTCTCCAAGAGTGCC 1395

51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSe 67
1396 AACTACTGCCATACGCTCTCAGGAGACCAATAGGCTTAATCCTGTTGGG 1445

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
1446 AGAAGTTCCTTGGAAACATGTATGAATCAAGCAGCCTTCACATATCA 1495

84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
1496 GCAAGTTACCAAGGCAAGCACAGTGTCAAGGTTTGGGCAAACTACC 1545

101 ProLeuGluSerGluPheValLysTrpArgAspValValProCy 117
1546 CTTGATCTCTCAGCTAACATAGT...CTGGATGCTGTAGACGTTCTCT 1592

117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
1593 TGGGACCGGATTTCATCTGGTGTGAATGACACCTCTCTACTATATACG 1642

134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
1643 AGTACATTGCTATGATATTGCTCAGGTAAATCTGAAGTATCTGCTGAAA 1692

151 ValArgPheHisHisLys 156
1693 CTGAATTCATTTTAAG 1710

```

seq_name: gb_pr:HUMPOLP

seq_documentation_block: 3640 bp mRNA 08-JAN-1995

LOCUS Human poly(ADP-ribose) polymerase mRNA, complete cds.

DEFINITION M18112

ACCESSION M18112

VERSION M18112.1 GI:190166

KEYWORDS polymerase.

SOURCE Homo SIV40 transformed fibroblast, cDNA to mRNA, clone pPAP.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3640)

AUTHORS Uchida,K., Morita,T., Sato,T., Ogura,T., Yamashita,R., Noguchi,S., Suzuki,H., Nunoya,H., Miwa,M. and Sugimura,T.

TITLE Nucleotide sequence of a full-length cDNA for human fibroblast poly(ADP-ribose) polymerase

JOURNAL Biochem. Biophys. Res. Commun. 148 (2), 617-622 (1987)

MEDLINE 88076933

COMMENT Draft entry and computer readable sequence for [1] kindly provided by K.Uchida, 02-MAR-1988.

FEATURES

Location/Qualifiers

1..3640

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="1q41-q42"

<1..3640

/note="POLP mRNA"

140..3184

/gene="ppol"

140..3184

/gene="ppol"

/note="poly(ADP-ribose) polymerase"

/codon_start=1

/db_xref="GDB:G00-119-508"

/protein_id="AAA60137.1"

/db_xref="GI:190167"

/translation="MAESSDKLYRVEYAKSGRASCCKSESIPKDSLRMALWQSPMF

DGKVPWHYHSCFWKRGHSIRHPDVQVDFSELWDDQCKVKRTAEAGVTGKGQDGI

GSKAEKTLGDFAAEYAKSNRSTCRGCMKEIKGQVRLSKMVDPEKPOLGMDIRWYHP

GCYKNEELGFPEYASQLKGFSLATEDKALKKQLPGVKSEGRKGDGVDE

VAKRKKKDKSKLEKALKAQNDLWIKNDKLVKVCSTNDLKLIFNKQVPSCE

SAILDRVADGMVFGALLPCRECSQGLVFKSDAYCYCTGDTAWTKCMVKTOTPNRKEW

TPKPEREISLYLKLKVKQDRIFFPETASVAATPPSTASAPAVNSSASADKPLSN

MKLTILGKLSRNKDEVRAMIEKLGKLTGTANKASLCISTKKEVERMKNKEEYKEAN

IRVSEDFLDVVSASTKSLQELFLAHLSPWCAEYKAEVPAAPRGSKGAALSKSK

QVKEGINKSEKRMKLTLLKGAADVPDPSGLSHSAHLEKGGKVSATLGLVDIVKGT

NSYKQLLEDKKNRYWIFRSGRWGTVIGSNKLEQMSKEDAEIHFMLYBEKTNAM

AWHKNFTYKPKFYPLEIDYGDDEAVKLTVNPGTKSLPKPQVQLIKMIFDVESM

KKAMVEYIDLOKPLGLSKRQIOAAYSILSEVOQAVSOGSSDQILDLNRFYTLI

PHDFGMMKPPLLNNADSVQAKVEMLDNLLDIEVAYSLRGSSDSDKDPIDVNEKLU

TDIKVVDROSEAEIIRKYVKNTHATTHNAYDLEVIDFKLERGEQORYKPKFOLHN

RRLWHSRTTNFAGLSQGLRIAPPEAPVTGYMFGIKGFADWVSKSANYCHPTSQD

PTGLILLGEVALGNMYELKHASHISLKPGRHSVKGKLTTPDPFSPANISLDGVDVPLG

TGISSGNDTSLLYNEYIIVDYAQNLYLLKFNFKTSLW"

BASE COUNT 999 a 833 c 1008 g 800 t

ORIGIN Chromosome 1p11-qter.

alignment_scores:

Quality: 517.50 Length: 156

Ratio: 3.777 Gaps: 1

Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:

US-09-236-995D-4 x HUMPOLP ..

Align seg 1/1 to: HUMPOLP from: 1 to: 3640

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17

1246 AACCGAAGATTGCTGTGGCAGCGGTCCAGGACCAACATTTTGTCTGGAT 2754

```

17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
2755 CCTGTCACAGGCTCTCGAGTAGCCGCCCTGAAGCGCCGCTGACAGCT 2804
34 yrMetPheGlyLeuGlyLeuPheAlaAspLeuValSerLysSerAla 50
2805 ACATGTTGGTAAAGGATGATTTCCTGCTGACATGGCTTCCAAAGAGTCC 2854
51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
2855 AACTACTCCATACGCTCTCAGGAGACCAATAGCTTAATCCTGTTGGG 2904
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysAlaThrSerMetA 84
2905 AGAAGTTCCTTGGAAACATGTTGAAGTGAAGCAGCTTACATATCA 2954
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
2955 GCAAGTTACCCAGGCGCAAGCAGACATGTCAGAGGTTTGGGCAAACTACC 3004
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
3005 CCTGATCCTTCAGCTAATAGT...CTGGATGCTGATAGCTTCTCT 3051
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
3052 TGGACCCGGGATTCATCTGCTGTGAATGACACCTCTCTACTATATAACG 3101
134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
3102 AGTACATGCTCTATGATATCTCAGGTAAATCTGAAGTATCTGCTGAAA 3151
151 ValArgPheHisLysLys 156
3152 CTGAATTCATATTTAAG 3169
seq_name: gb_pat:A52134

```

```

seq_documentation_block:
LOCUS A52134 3792 bp DNA PAT 11-MAR-1997
DEFINITION Sequence 1 from Patent WO9618737.
ACCESSION A52134
VERSION A52134.1 GI:2304739
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 3792)
AUTHORS Buerkle,A., Zur,H.H. and Kuepper,J.
TITLE VECTORS AND VIRUSES FOR USE IN GENE THERAPY
JOURNAL Patent: WO 9618737-A 1 20-JUN-1996;
DEUTSCHES KREBSFORSCH (DE)
FEATURES
source
1..3792
location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
96..3137
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:2304740"
/translation="NAESDRLRYVEYAKSERASCKSCSEIPKDSLRMAINQSPMF
DKVPHVHFSCFVGHSHIRHPDVEVDFGSELRWDQKVKKTAEGAGVTCKGQDGI
GSAKXTLGDFAEYAKNSHSTCKGMEKIEGQVRLSKKMYDPEKPOLGMDLRVHP
GCFVKNRESELGRPEYSASQLKGFSLIATEDKEALKQLPGVKSEGRKRGKQVGDVE
VAKKSKKEDKDLKLEKQNDLIWNIDKELKVCSTNDLKELLIFNKQOVPSGE
SAILDRVADGVFGLLPCECGSLVFKSDAYCTGDVMTKCMVKVQTPTNRKEW
TPKEPEISYLNKKLVKKODRIFPPETSASVAHPPPTASPAANVSASADKPLSKM
KILTGLDNKKDEVKAMLEKLGKLTGTANKASLCISTKKEVKMKMKMEVKEANI
RVVSEDFQDSASTKSLQELFLAHLSPWAEVRAEPEVVPAPRGKGAALSASKSG
QYKEGINKSEKRMKLTLEKGAADVPDGLSEHSAHLEKGVKVFATGLGLDIVKGIN
SYVKLQOLDDDDKENRWIPRSWGRVGTIGSNKLEQMPFSKEDAIIEHFMKLYBEKTGNA

```

```

WHSKNTFYPKKFFPLEIDYCODEEAVKLTVPNTKSKLPKVQDLIKMLFDVESMK
KAMVEYEDLOKMPGLKSKRQIOAAISILSEVOQAVSQSSQSSQIIDLNSRPTLIP
HDFGKKPPLNNADSVQAKVEMLDNLLDIEVAYSLRGGSDSSKDPIDVNEKLLK
DIKVYDRSEAEAIIRKYVNTHTATPNAVDLEIDIFKIEREGECORYKPFQOLHNR
RLWHGSRRTTFAGILSQGLRIAPPEAPVTGMEFGKGYFADMVSKSANTYCHTSQSDP
IGLILGEVALGNMVELKHASHISKLPKXHSVKGLGKTTTPDPSANISLDGVDVPLGT
GISSGVNDTSLLYNEYIVYDIAOVNLAKYLKLNFKFTSLW"
BASE COUNT 1048 a 846 c 1035 g 863 t
ORIGIN

```

```

alignment_scores:
Quality: 517.50 Length: 156
Ratio: 3.777 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179
alignment_block:
US-09-236-995D-4 x A52134
Align seg 1/1 to: A52134 from: 1 to: 3792

```

```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyTl 17
2658 AACCGAAGATGCTGTGGCAGCGGTCCAGGACCAACCACTTGTCTGGAT 2707
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
2708 CCTGTCCAGGCTCTCGGATAGCCCGCTGAAGCGCCGCTGACAGGCT 2757
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
2758 ACATGTTGGTAAAGGATCTATTCGCTGACATGCTTCCAAAGAGTGGC 2807
51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
2808 AACTACTCCATACGCTCTCAGGAGACCAATAGGCTTAATCCTGTTGG 2857
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
2858 AGAAGTTCCTTGGAAACATGTTGAAGTGAAGCAGCTTCCATATCA 2907
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
2908 GCAAGTTACCCAGGCGCAAGCAGTGTCAAAGGTTTGGGCAAACTACC 2957
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
2958 CTGATCCTTCAGCTAATAGT...CTGGATGCTGATAGCTTCTCTCT 3004
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
3005 TGGACCCGGGATTCATCTGCTGTGAATGACACCTCTCTACTATATAACG 3054
134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
3055 AGTACATGCTCTATGATATTCGCTCAGGTAAATCTGAAGTATCTGCTGAAA 3104
151 ValArgPheHisLysLys 156
3105 CTGAATTCATATTTAAG 3122
seq_name: gb_pr:HUMRISDAD

```

```

seq_documentation_block:
LOCUS HUMRISDAD 3795 bp mRNA PRI 09-JAN-1995
DEFINITION Human poly(ADP-ribose) synthetase mRNA, complete cds.
ACCESSION J03473
VERSION J03473.1 GI:337423
KEYWORDS poly(ADP-ribose) synthetase.
Human placenta, cDNA to mRNA (library of H. Okayama), clones
PPARS[1,11,21,32,32,41,-F].
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

/protein_id="AAF61293.1"
/db_xref="GI-7363322"
/translation="MKLTGGAADPDGSLHSAHVLEKGGKVFSAITGLVDIVKGT
NMYKIQLEDDRESYWIIFRSRGVTGIVGNSKQMPSEKDAVEHFMKLYEEKTGN
ASHSKDFTRYPKFYLEIDYGDDEAVKLTVPKGTSKLPVQELVGMIFDVESM
KRALVEIDIDKMPLEKLSRRIOQAAYSITLSEVPVSGSSESILDLNSRFYTLI
PHDFGMKKPPLNNADSVLAKVEMLDNLLDIEVAYSLRGGSDSKDPIDVNYEKLK
TDIKVVDROSEAEVIRKYVKNTHATHNAYDLEVIDIEKIEREESORYKPFRLQHN
RRLWHSRTTNFAGILSGLRIAPPEAPVTGFMFGKGYFADMVSKSANYCHTSOGD
PGLIMGEVALGNMYELKHASHISKLPKHSVKGKLTTPDPSASITLEGVEVPLG
TGTPSVNDTCLLYNEYIYVDAQVNLKLLKFNFKTSLW"
3'UTR
BASE COUNT 472 a 424 c 493 g 343 t
ORIGIN

```

```

alignment_scores:
  Quality: 516.50      Length: 156
  Ratio: 3.798        Gaps: 1
Percent Similarity: 87.179 Percent Identity: 61.538

```

```

alignment_block:
US-09-236-995D-4 x AFI26717

```

Align seg 1/1 to: AFI26717 from: 1 to: 1732

```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPhValGlyI1 17
|||||
1251 AACCGGAGGCTGCTGTGGCAGGCTCCAGGACCACCAACTTTGTGTCAT 1300

17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||
1301 CCTGTCGACGGTCTGCGGATAGCCACCTGAAGCGGCTGTGACAGGCT 1350

34 yrMetPheGlyLysGlyLeuTrpPheAlaAspLeuValSerLysSerAla 50
|||||
1351 ACATGTTGGGAAGGATCTACTTTCGCGACATGGTGTCCAAAAGTGCA 1400

51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
1401 AACTACTGCCACATCTCAGGAGACCGGATTGGCTTAATAATGCTGGG 1450

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||
1451 AGAGTTCCTTGGAAATATGTATGAACCAAGCATGCTTCACATATCA 1500

84 spLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
1501 GCAAGTACCAAGGCAAGCAAGTGTCAAGGTTGGGAAAACCAACC 1550

101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||
1551 CTGACCCCTTCGGCCAGCATCACC...CTGGAGGGTGTAGAGGTTCCACT 1597

117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||
1598 GGAACAGGGATCCATCTGTGTGTCAACAGCACCTGCTGCTGTATATG 1647

134 LuTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
|||||
1648 AGTACATGTTCTAGACATGCTCCTCAGGTGATCTCAAAATCCTGCTGAAA 1697

151 ValArgPheHisLysLys 156
|||||
1698 CTCAAGTTCATTTTAAG 1715

```

seq_name: gb_pl.AC006593

```

seq_documentation_block:
LOCUS AC006593 79663 bp DNA PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 176 of 255 of the
complete sequence. Sequence from clones F16d14, T28P16.
ACCESSION AC006593 AB002093

```

```

AC006593.4 GI:6598616
HTG.
SOURCE
ORGANISM

```

```

Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 79663)
AUTHORS
Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
Fuji,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
Feldblum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Unayam,L.,
Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
Venter,J.C.

```

```

TITLE
Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
JOURNAL
Nature 402 (6763), 761-768 (1999)
MEDLINE
20083487
PUBMED
10617197
REFERENCE
2 (bases 1 to 79663)
AUTHORS
Lin,X.

```

```

Direct Submission
JOURNAL
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT
On Dec 17, 1999 this sequence version replaced gi:4432811.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at/at.html).

```

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6223, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```

FEATURES
    source
        1..79663
            /organism="Arabidopsis thaliana"
            /cultivar="Columbia"
            /db_xref="taxon:3702"
            /chromosome="II"
    misc_feature
        1..79462

```

```

/note="Sequence from clone F16D14"
complement(join(<1349..1687,1764..1967,2138..2263,
2342..2623,2709..2879,2963..3124,3898..3961,4063..>4406))
/gene="At2g31170"
complement(<1349..>4406)
/gene="At2g31170"
/note="F16D14.1"
complement(join(1349..1687,1764..1967,2138..2263,
2342..2623,2709..2879,2963..3124,3898..3961,4063..4406))
/gene="At2g31170"
/codon_start=1
/product="putative cysteinyl-tRNA synthetase"
/protein_id="AAD20662.1"
/db_xref="GI:4432812"
/translators="MASSVLEKSCRPETPIRFSSLPKSPQIOPPLRGKETOLRR
CFTLLSLDGGAPISGGKELWHLNTHMSKKELFKPKVEKGMVYCGVAYDLSHIG
HARVYVTFLLRLHLYGVEYSVRNFTDVKIIRAKALEEDPISRRPCEFN
RMEQQLDPSQVPRVSHIPQIIDLKQILDNGYAYKVDGDIYFSDKFPPTYKGLS
GRLEONRAGERVAVTRKKHPADFAWKAKEGEPWPMGRPGWHIECSAMSA
AYLGYSPIHGGMDLVFPHHENIAQSCAACDSNYSWIHNGFVTVDSKMSKSLG
NFTIRQVIDLYHPLALRFLMGTHVRSPIYSDFLLESASERFIYOTLHDCESAL
GEKDSFENSVSFSDILTSINTERTFEVASMSDDLITPYLAMSEPLKTINDLIHTR
KKQKARRESLKALETIRDLVLTILGLMPTSYSEVLEQLKALKARAGLKEDVLQR
VQERTDARKKEYERSDAIRKDLAKVGIALMDSPEGTWRPAIPLAQEPVTTTP"
complement(join(5525..6082,6603..6732,6833..>6965))
/gene="At2g31180"
complement(5525..>6965)
/gene="At2g31180"
/note="F16D14.2"
complement(join(5596..6082,6603..6732,6833..6965))
/gene="At2g31180"
/codon_start=1
/product="putative MYB family transcription factor"
/protein_id="AAD20663.1"
/db_xref="GI:4432813"
/translators="MGRAPCCERKMGVKGWPTPEEQILINYLHLYGHNSWRALPKHA
GLRCGRLRWNYLRLPDIKRGNTQPEEQIILHESLGNRWSAIAKLPGRTDN
EIKNVHHLKRLKSLNNGDQKDVINETNEDKGSVIDTASLOOFSNITFE
DISNDKNDJMSVEDISALJDDSFSDVSDVSNKNEKKIEDWGLIDRNSKKCSYS
NSKLYNDMEFEDVFTSNRRIEFSDIPEF"
complement(7897..7920)
/rpt_family="Poly_A"
8038..8151
/rpt_family="(TA)A)n"
9946..9977
/rpt_family="(TA)n"
complement(join(<10248..10673,10764..10886,10983..11162,
11326..11423,11506..11599,11690..11749,11871..11955,
12057..>12145))
/gene="At2g31190"
complement(<10248..>12145)
/gene="At2g31190"
/note="F16D14.3"
complement(join(10248..10673,10764..10886,10983..11162,
11326..11423,11506..11599,11690..11749,11871..11955,
12057..12145))
/gene="At2g31190"
/note="unknown protein"
/codon_start=1
/protein_id="AAD20664.1"
/db_xref="GI:4432814"
/translators="MKVVDDARVPVKWVESFLNKFFPGYPYSVNEGILRYTQFRAL
QHPSSAALSVLSQSLFLAAGLRPTPAQTAVSVKLLKDCGMHVKGLCSNLGMRDSE
PKRWRLADVLIDGLTGLVSLPCLPHLFLEMAGLGNFAKGMATVAARATRLPIYSF
AKENLSIDIFAKGATSLFNAGIGAGIOLASTICSSMEGLKLVGSLISLVHVISV
EOMGRVINTLPORALIVANFKTGKVPSPDLRFQEDLMFPERPIDAGNVKGR
KLVKAVPSEVKQVFEVEKFLLSHGSWTDWLEHDAITGEDALRGVLVAAYKSM
TKIYNRPDDIILODAYDKMDFNPFLSQVQAKGWYTDRELDGTGTRFAM"
join(12936..12984,13233..13498,13588..13809)
/gene="At2g31200"
12936..13809
/gene="At2g31200"
/note="F16D14.4"

```

```

CDS
join(13251..13498,13588..13738)
/gene="At2g31200"
/codon_start=1
/product="actin depolymerizing factor 6"
/protein_id="AAD20665.1"
/db_xref="GI:4432815"
/translators="MGVADESKTTFLELQKKTHRVVVFVKIDESKKEVVVEKTNPTE
SYDDPLASLPNDCRYAYVDFDFTSENCQSKIFFFAWSPSTGIRAKVLYSTKQD
LSRELUQTHYEIQATDPTVEVDLEVRAN"
join(<15445..16170,16266..>16739)
/gene="At2g31210"
<15445..>16739
/gene="At2g31210"
/note="F16D14.5; predicted by genefinder"
join(15445..16170,16266..16739)
/gene="At2g31210"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD20666.1"
/db_xref="GI:4432816"
/translators="MYEBSGCFDPNSMVDNNGGCAAEETFTVSHQFQPPILGTTNSP
DDDLKPTMDESVPSVLSIPNSETONNNTNNHLLNOMIQESNWCVSNDNFMFM
NTHSHNTTTPDLLSLHLPRCSMSLPSSDINAGSCFTYDPLFLNLPQPLIPS
NDYSYLIGDITNTTTQDESNGVDENNAQFDGSIIEFSKEIRKGRGKKNKPFYTT
ERERCHLNERYEALKLITPSPKGDRAILOGDIDIINELRRVSELKYLVRKRCG
GRHKNEVDNKNKNDHGNEDDDDDENKPKPSDVIDQCSNNSLRCSWLQRK
SKTVEVDVRIVDDEVTIKVVQKKINCLLVSKVLQDLQLDLHHVAGQIGEHYSELF
NTKVSQD"
complement(16419..16456)
/rpt_family="(CAT)n"
20630..20718
/rpt_family="(TAAAA)n"
join(21922..22723,22818..23267,23350..>23451)
/gene="At2g31220"
21922..>23451
/gene="At2g31220"
/note="F16D14.6"
join(21932..22723,22818..23267,23350..23451)
/gene="At2g31220"
/note="unknown protein"
/codon_start=1
/protein_id="AAD20667.1"
/db_xref="GI:4432817"
/translators="MGCDFDPTPAEVTVESSFSQAEPPLPPQVLVAGSTNSNCSVE
VEELSEFLSPQDCPQASSTPLQFHINPPPPPPDQLHNNLIHQMASHQOHSNWD
NGYQDFHLSPNSATTPDLLSLHLPRCSLPNHPHPSMLPTSFSDINSSSSAAVMI
DPLFLNPMOPRODQNLNGLGVEDQIQMDANGMVLYFGANNNGGPNELID
LEFNNGVTKGRGSRKTSPTERRRVHFNDRFDLKNLIPNPKIDRASIVGEAID
YIKELLRTIEEFKMLVKRCGRFRKRKRARGGGGDEDEEDTVNYKPSQVEDQ
SCFNKNNNNSLRCSWLKRKSKVTEVDVRIIDDEVTIKLVQKKINCLLFTTKVLQDLQ

```

alignment_scores:

Quality: 512.50 Length: 232
Ratio: 3.768 Gaps: 4
Percent Similarity: 58.621 Percent Identity: 49.138

alignment_block:

US-09-236-995D-4 x AC006593/rev ...

Align seg 1/1 to reverse of: AC006593 from: 1 to: 79663

2 LysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyIleLe 18
||||: ||| ||||||||||||||||||||||||||||
73774 AAATAATCTGCTTTCAGGTTCTCGATTACGAATTTTGTGAATATT 73725

18 userGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyTyr. 34
|:|||||||||||||||||||||||||||||
73724 GAACCAAGGACTGAGAAATTGCACCTCCAGAAGCTCCTGCTACTGGTTTACA 73675

34 34

73674 TGGTCAGTACTCTATCTGTCTTATTGTTGGGGATCTTATCAATTGAGA 73625

```

34 ..... 34
73624 GTTCTTCACTACTAGTTAACTGTTGAGAGAACTGGAACCTTTTAAAC 73575
35 .....Met...PheGlyLysGlyLeuTyrPheAlaAspLeuValSerLy 48
73574 CCCATGTTTTCAGTTTGGAAAAGGATATACCTTGTGACCTTGTGAGTAA 73525
48 sSerAlaGlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetL 65
73524 AAGTCTCTCAGTACTGCTACACTGTGAAGAAAAATCCGGTGGTCTAATGC 73475
65 euLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThr 81
73474 TTCTGAGTGAAGTTGCAATTTGGGAGAAATACATGAGCTAACAAAAAGCTAA 73425
81 ..... 81
73424 GTGAGCCCATATATACAGTTTTCAGGTTATGCTACTAACCAATTTGGTTCT 73375
81 ..... 81
73374 TATCTTGAAACTATATGAACAACCTTAAGATATTACTTACCTCATATAT 73325
82 .....SerMetAspLysProProArg 88
73324 GACTTCTCTTATGTCCTCAATCTTTAGTATATATGGAATAACCTCCGAGA 73275
89 GlyLysHisSerThrLysGlyLeuGlyLysThrValProLeuGluSerG 105
73274 GGGAAACACTCGAACCAAGGCTCGGCAAGAAAGTGCTCAAGATTCCGA 73225
105 uPheValLysTrpArgAspValAlaValProCysGlyLysProValP 122
73224 GTTGGCAAGTGGAGAGGTGATGTGACTGTTCCCTGTGGAAACCTGTT 73175
122 roSerSerIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyr 138
73174 CATCAAGGTCAGGCTCTGAGCTTATGTACAAATGATATATGCTCTAC 73125
139 AsnThrSerGlnVal.....LysMetGlnPheLeuLeu 149
73124 GATACAGCCCAAGTAACTGTGTTTCCAAATAATCTTCTCTGTT 73080

```

seq_name: gb_ov.XLPARPG

```

seq_documentation_block:
LOCUS      XLPARPG      3617 bp      mRNA      VRT      18-JAN-1993
DEFINITION X.laavis PARP gene encoding poly(ADP-ribose) polymerase.
ACCESSION  Z12139
VERSION    Z12139.1 GI:64967
KEYWORDS   poly(ADP-ribose) polymerase.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
            Xenopodinae; Xenopus.
            1 (bases 1 to 3617)
            Saulier-le Drean,B.M.
            Thesis (1992) Lab. de Biol. et Genet. du Developpement, Universite
            de Rennes I. URA CNRS 256
            2 (bases 1 to 3617)
            Saulier-le Drean,B.M.
            Direct Submission
            Submitted (15-MAY-1992) Saulier-le Drean B.M., Lab. de Biol. et
            Genet. du Developpement, Universite de Rennes I. URA CNRS 256
            Campus de Beaulieu, Av. du Gal Leclerc, Rennes-cedex, FRANCE, 35042
FEATURES             Location/Qualifiers
     source           1..3617
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /germline
                     /sex="female"

```

```

/dev_stage="Adult"
/tissue_type="Ovary"
/cell_type="Oocyte"
/clone_lib="oligo dt primed lambda gt10 library of D.
MELTON"
/clone="clone parp 923"
1..2997
/gene="PARP"
/gene="PARP"
1..2997
/gene="PARP"
/EC_number="2.4.2.30"
/function="poly ADP-ribosylation, DNA repair"
/notes="protein sequence is in conflict with the conceptual
translation: poly(ADP-ribose) polymerase"
/citation=[1]
/codon_start=1
/product="NAD(+) ADP-ribosyltransferase"
/protein_id="CAA78126.1"
/db_xref="GI:1334661"
/db_xref="SWISS-PROT:P31669"
/translation="AKSRASCKKCGDNIAKESGLAIWQSPMFDCGVPHWHYSCF
WKARVLSQDIYGYTELWEDQEMIKAIETGGAAGAGDSKGGKGMETLNDFAAE
YAKNSRACKGEQIEKQIRISKSVDPOLGMIDRWHPDCEVSSRELDLFLP
SYSASQLKQFTILSAEDKSLKLPVAKNEGRKADEVDGHSAAATKKIKKKEKES
KLERLLKEQTELIWHIKDELKVCSTNDLKELLIANKQOVPSGETNIVDVSGMAFG
ALLPCECSGQFVFKDAYCTGDSAWTKCAKTOPTNRKDMWTPKEFHEIPLYKFF
KFKRDRAPFCAAPTPLSPAPEPKTVEETPECKPLNTKYLILGLSKNKDEY
KTLLEGIGKVAGSAHKANICISINKEVKKMSKMEVKANRVVSDDFLKEVESGK
SVQELSOFGISSWGAIEKQEAQVQTEKQSPSSGVPAGSKGKVEKSGSKSKMKML
TVRGAAIDPDSELEDSHLETKGKIFSATGLVDITRTGNSYFKLLEHDDRSRY
WFRSGRQGVTVIGSKLEMSKDEAIEHFLNLYQDKTGNWHSNFTKYPKFPYPL
EIDYQGEVGVKLSVGAGTKSLAKPQBELIKLIPVESMKAMVFEIDLOKMPLG
KLSKROIQSAYSILSOVQAVSESLSEARLILDSNOFTLIPHDFGKKKPELNNLEY
IQAKVQMLNDLIEAVYSLRGAGDGEKDPIDYKYEIKTDIKVVKAKDSESRILIC
DIVANHTADHINADLELIFKIDREGEYQRYKPKOLHNRLQLLWHSGSTTFAGIL
SQGLRIAPPEAPPTGYMFGKGIYFADWVSXSNYCHAMGPSPIGLILGSEVALGNHE
LKAASQITKLPKGVKSGVGTAPDPSATVQDGVDPVPLGKGTSANISDTSLLYNEY
IYVDIAQVNLKLLKLFNYKGGMMW"
3'UTR
polyA_signal
polyA_site
BASE COUNT 1131 a 757 c 910 g 819 t
ORIGIN

alignment_scores:
Quality: 511.50 Length: 156
Ratio: 3.734 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:
US-09-236-995D-4 x XLPARPG ..
Align seg 1/1 to: XLPARPG from: 1 to: 3617
1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyLyl 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2515 AACCGCCAGCTGCTTTGGCAGCGCTCCCGCACCAAAATTTTGCAGGAAT 2564
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2565 ATTGTCTCAGGCTCTCCGAATTCCTCCGCCAGAGCTCTCTGTACCGGCT 2614
34 YrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2615 ATATGTTTGGCAAGAGTATCTATTTTTCGCCAGATGTTATCCAAAGATGCA 2664
51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2665 AACTACTGTGATGCAATGCCGGGTAGCCCCATAGGCTGATCTTACTGGG 2714
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```
2715 GGAGGTGGCCCTCGGAACATGCATGAGCTGAAAGCAGCATCACAAATTA 2764
84 splysProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
   ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2765 CAAAACTTCCCAAGGCAACACAGTGTAAAGGTTTGGGTAGAACTGCA 2814
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
   .. ||| ||| |||:||||| ||| ||| ||| ||| ||| ||| |||
2815 CCAGATCCCTCAGCTACTGTACAG...CTGGATGGAGTGGATGTTCCCTCT 2861
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
   ||||| :|| :||:||||| :||:||||| :||:||||| :||:|||||
2862 CGGAAAGGAAGTCTCTGCAATATCAGTCACACACAGCCCTGTTGTATATG 2911
134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
   ||||| ||||| ||||| :||:||||| :||:||||| :||:|||||
2912 AATATATTGTGTATGATATTGCTCAGGTCAACCTGAAGTACCTGCTGAAG 2961
151 ValArgPheHisLys 156
   :||:|||||:|||||
2962 CTCAGTTCAACTACAAA 2979
```


OM of: US-09-236-995D-4 to: N_Geneseq_1101:* out_format : pfs
 Date: Mar 9, 2002 8:30 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
 -Q/cg2_1/USPTO.spool/US09236995/runat_08032002_113149_25135/app_query.fasta_1.1273
 -DB=N_Geneseq_1101 -QMT=fastap -SUFFIX=mg -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOFCI=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM-ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09236995 -CGN1_1_309 -NCFU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-236-995D-4
 Query length: 157
 Database: N_Geneseq_1101:*
 Database sequences: 930621
 Database length: 428662619
 Search time (sec): 250.160000

score_list:

| Sequence | Strd Orig | 2Score | EScore | Len | Documentation |
|--|-----------|--------|---------|---------|---------------|
| /SIDS2/gcgdata/geneseq/NA1999.DAT:AAF89542 | + | 828.00 | 1857.54 | 2.2e-95 | 2949 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF60615 | + | 828.00 | 1856.60 | 2.5e-95 | 3211 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF60618 | + | 828.00 | 1856.59 | 2.5e-95 | 3212 |
| /SIDS2/gcgdata/geneseq/NA1999.DAT:AAF23799 | + | 517.50 | 1144.26 | 1.2e-55 | 3580 |
| /SIDS2/gcgdata/geneseq/NA1996.DAT:AAF13732 | + | 517.50 | 1143.62 | 1.3e-55 | 3793 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63954 | + | 500.50 | 1107.11 | 1.4e-53 | 3045 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF60617 | + | 496.50 | 1101.82 | 2.7e-53 | 2147 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF60616 | + | 453.50 | 1002.60 | 9.1e-48 | 2295 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF5996 | + | 441.50 | 985.91 | 7.7e-47 | 864 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF5996 | + | 441.50 | 979.34 | 1.8e-46 | 1566 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF5996 | + | 441.50 | 977.71 | 2.2e-46 | 1814 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF44287 | + | 441.50 | 977.54 | 2.3e-46 | 1843 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF82090 | + | 441.50 | 977.54 | 2.3e-46 | 1843 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF77118 | + | 441.50 | 977.13 | 2.4e-46 | 1912 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF85341 | + | 441.50 | 971.44 | 4.9e-46 | 3200 |
| /SIDS2/gcgdata/geneseq/NA1995.DAT:AAF04321 | + | 435.50 | 952.03 | 5.9e-45 | 5345 |
| /SIDS2/gcgdata/geneseq/NA1995.DAT:AAF04321 | + | 429.50 | 945.91 | 1.3e-44 | 2682 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAH14941 | + | 428.00 | 945.83 | 1.3e-44 | 1980 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ80132 | + | 358.50 | 799.94 | 1.8e-36 | 595 |
| /SIDS2/gcgdata/geneseq/NA1995.DAT:AAZ04220 | + | 350.50 | 770.74 | 7.5e-35 | 1592 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44291 | + | 227.50 | 489.07 | 3.6e-19 | 1587 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44290 | + | 227.50 | 488.05 | 4.1e-19 | 1740 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44288 | + | 225.50 | 480.56 | 1.1e-18 | 2265 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44289 | + | 225.50 | 480.56 | 1.1e-18 | 2265 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ45690 | + | 173.00 | 350.53 | 1.9e-11 | 5490 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ92076 | + | 173.00 | 350.53 | 1.9e-11 | 5490 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85304 | + | 133.50 | 287.19 | 6.4e-08 | 472 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63955 | + | 127.00 | 245.19 | 1.4e-05 | 5482 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ66826 | + | 123.00 | 239.56 | 2.9e-05 | 3984 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ29627 | + | 123.00 | 239.15 | 3.0e-05 | 4134 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ29632 | + | 123.00 | 238.24 | 3.4e-05 | 4491 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ29633 | + | 123.00 | 237.83 | 3.6e-05 | 4657 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63875 | + | 118.00 | 249.37 | 8.2e-06 | 582 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ77578 | + | 118.00 | 240.12 | 2.7e-05 | 1344 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63919 | + | 118.00 | 231.35 | 8.2e-05 | 2971 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85295 | + | 118.00 | 230.35 | 9.4e-05 | 3254 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63990 | + | 118.00 | 230.16 | 9.6e-05 | 3308 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63917 | + | 118.00 | 230.02 | 9.8e-05 | 3353 |
| /SIDS2/gcgdata/geneseq/NA1999.DAT:AAZ23566 | + | 118.00 | 229.86 | 0.00010 | 3400 |

/SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63837 + 118.00 229.52 0.0001 3508
 /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ91487 + 118.00 228.64 0.0001 3797
 /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63918 + 118.00 228.63 0.0001 3799
 /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ91488 + 118.00 228.59 0.0001 3816
 /SIDS2/gcgdata/geneseq/NA2001.DAT:AAF63926 + 118.00 227.72 0.0001 4127

seq_name: /SIDS2/gcgdata/geneseq/NA1999.DAT:AAF89542

seq_documentation_block:

ID AAX89542 standard; cDNA; 2949 BP.

AC AAX89542;

DT 06-OCT-1999 (first entry)

DE Maize poly ADP-ribose polymerase gene.

KW PARP; poly ADP-ribose polymerase; metabolism; maize; transform; ss.

OS Zea mays.

PH Key
 FT CDS
 FT Location/Qualifiers

1..2949
 /*tag= a
 /product= "PARP"
 /note= "Poly ADP-ribose polymerase"
 /transl_except= (pos:1453..1455, aa:Xaa)
 /transl_except= (pos:1495..1497, aa:Xaa)
 /transl_except= (pos:1504..1506, aa:Xaa)
 /transl_except= (pos:1534..1536, aa:Xaa)
 /transl_except= (pos:1537..1539, aa:Xaa)
 /transl_except= (pos:1561..1563, aa:Xaa)
 /transl_except= (pos:1582..1584, aa:Xaa)
 /transl_except= (pos:1588..1590, aa:Xaa)
 /transl_except= (pos:1594..1596, aa:Xaa)
 /transl_except= (pos:1975..1977, aa:Xaa)
 /transl_except= (pos:2011..2013, aa:Xaa)
 /transl_except= (pos:2014..2016, aa:Xaa)
 /transl_except= (pos:2026..2028, aa:Xaa)
 /transl_except= (pos:2047..2049, aa:Xaa)
 /transl_except= (pos:2050..2052, aa:Xaa)
 /transl_except= (pos:2053..2055, aa:Xaa)
 /transl_except= (pos:2056..2058, aa:Xaa)
 /transl_except= (pos:2077..2079, aa:Xaa)
 /transl_except= (pos:2095..2097, aa:Xaa)
 /transl_except= (pos:2107..2109, aa:Xaa)
 /transl_except= (pos:2116..2118, aa:Xaa)
 /transl_except= (pos:2119..2121, aa:Xaa)
 /transl_except= (pos:2122..2124, aa:Xaa)
 /transl_except= (pos:2125..2127, aa:Xaa)
 /transl_except= (pos:2128..2130, aa:Xaa)
 /note= "Xaa = unknown"

WO9937789-A1.

29-JUL-1999.

26-JAN-1999; 99WO-US01591.

27-JAN-1998; 98US-0072785.

(PION-) PIONEER HI-BRED INT INC.

Mahajan P, Zuo Z;

WPI: 1999-444613/37.

P-PSDB; AAY28464.

New maize poly ADP-ribose polymerase gene useful in transforming plants to alter their metabolic state
 Claim 1; Page 30-35; 48pp; English.

CC The sequence is the maize poly ADP-ribose polymerase (PARP) gene. The
CC sequence codes a PARP with 982 amino acids (AA28464). PARP is required
CC in the cell in most cases of DNA repair, recombination, rearrangement
CC and transposition. PARP gene and antisense gene can be used to transform
CC plant cells and alter the metabolic state of the transformed cell. This
CC is useful in enhancing disease resistance in plants and methods of
CC genetic transformation of plants. Plants transformed with either a sense
CC or antisense PARP nucleotide sequence may be utilized to increase
CC transformation frequency in plant cells. The enzyme also plays a role in
CC cellular stress, so may be beneficial for prevention of plant disease or
CC pathogen attack.
XX
SQ Sequence 2949 BP; 895 A; 539 C; 732 G; 756 T; 27 other;

alignment_scores:
Quality: 828.00 Length: 157
Ratio: 5.274 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-236-995D-4 x AAX89542 ..

Align seg 1/1 to: AAX89542 from: 1 to: 2949

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
2476 AACAGATGCTATTATGGCACCGGTTCAAGGTTGACGAATTTGTGGGAAT 2525
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
2526 TCTTAGTCAAGGGCTAAGAAATTTGCACCTCTCTGAGGCACCTGTTACTGGCT 2575
34 yrMetPheGlyLysGlyLeuTrpPheAlaAspLeuValSerLysSerAla 50
2576 ATATGTTTCGCAAGGCCCTCTACTTTGCAGATCTAGTAAGCAAGAGCGCA 2625
51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSe 67
2626 CAATACCTGTTATGTGGATAGGAATATCTCTGAGGTTTGTATGCTTCTTTC 2675
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
2676 TGAGGTTGCTTTAGGAGACATGTATGAACATAAGAAAGCCACGTCATGG 2725
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
2726 ACAACCTCCAGAGGGAAGCATTCGACCAAGGGATTAGGCAAAACCCGTG 2775
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProcy 117
2776 CCACCTGGAGTCAGAGTTTGTGAAGTGGAGGGATGATGCTAGTTCCTCG 2825
117 sGlyLysProValProSerSerSileArgSerGluLeuMetTyrAsnG 134
2826 CGCAAGCCGGTGCCATCATCAATTAGGAGCTCTGAACATCATGTACAATG 2875
134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
2876 AGTACATCGTCTACACACATCCAGGTGAAGATGCAGTTCCTGCTGAAG 2925
151 ValArgPheHisLysArg 157
2926 GTGCGTTTCATCACAGAGG 2946

seq_name: /SID92/gcgdata/geneseq/geneseq/NA2000.DAT:AA260615

seq_documentation_block:

ID AA260615 standard; DNA; 3211 BP.
XX
AC AA260615;
XX
DT 16-MAY-2000 (first entry)

XX
DE DNA encoding the poly(ADP-ribose) polymerase ZAP1 protein of Zea mays.
XX
KW ZAP1; poly(ADP-ribose) polymerase; PARP; poly(ADP-ribose) transferase;
KW programmed cell death; apoptosis; growth rate; stress; cold; pathogen;
KW pest; drought; heat; fungi; nematode; seed-shatter; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 113..3022
FT /*tag= a
FT /product= "ZAP1 protein"
XX
PN WO2000041173-A1.
XX
XX 27-JAN-2000.
PD
XX
PF 12-JUL-1999; 99WO-EP04940.
XX
PR 17-JUL-1998; 98US-0118276.
XX
XX (PLBZ) PLANT GENETIC SYSTEMS NV.
PA
XX Babyichuk E, Kushnir S, De Block M;
PI
XX
XX WPI: 2000-182436/16.
DR P-PSDB; AAY68833.
XX
XX Modulating cell death, growth and stress resistance in eukaryotes,
PT specifically plants, used, e.g. to impart fungus or nematode resistance
PT
XX
PS Example 1; Page 79-84; 126pp; English.
XX
CC The present sequence encodes the ZAP1 protein of Zea mays. This protein
CC is a poly(ADP-ribose) polymerase (PARP) protein (also known as
CC poly(ADP-ribose) transferase). The PARP enzyme is involved in programmed
CC cell death or apoptosis, and is a nuclear enzyme. The ZAP1 polynucleotide
CC sequences can be used for modulation of programmed cell death in
CC eukaryotic cells. The method is used, specifically in plants, to induce,
CC or protect against, programmed cell death, depending on the extent to
CC which PARP activity is reduced. Reducing expression of endogenous ZAP
CC class PARP only is also used to modulate programmed cell death, to
CC increase growth rate and to produce plant cells that are more tolerant
CC of stress (cold, chemical treatments, pathogens, pests, drought, heat,
CC etc., or during transformation). Particular applications are generation
CC of plants that are resistant to fungi or nematodes; are male or female
CC sterile; or have better seed-shatter properties. The methods are also
CC used to improve growth of transformed plant cells (and derived calli or
CC complete plants).
XX
SQ Sequence 3211 BP; 968 A; 604 C; 813 G; 826 T; 0 other;

alignment_scores:

Quality: 828.00 Length: 157
Ratio: 5.274 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-236-995D-4 x AA260615 ..

Align seg 1/1 to: AA260615 from: 1 to: 3211

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||
2549 AACAGATGCTATTATGGCACCGGTTCAAGGTTGACGAATTTGTGGGAAT 2598
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||
2599 TCTTAGTCAAGGGCTAAGAAATTTGCACCTCTCTGAGGCACCTGTTACTGGCT 2648

```

34 yrmMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
2649 ATATGTCGCGCAAGGCTCTACTTTGCAGATCTAGTAAGCAAGAGCGCA 2698

51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
2699 CAATCTGTTATGTGGATAGGAATAATCCCTGATAGTTTGATGCTTCTTC 2748

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||
2749 TGAGTTGCTTTAGGAGACATGTATGAACATAACAAAGCCAGCTCCATGG 2798

84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
2799 ACAAACTCCCAAGAGGGAAGCATTCGACCAAGGATTTAGCAAAACCGTG 2848

101 ProLeuGluSerGluPheValLysTyrArgAspValValValProCy 117
|||||
2849 CCATGTGAGTCAGAGTTTGGAAGTGGAGGATGATGCTGTAGTTCCTG 2898

117 sGlyLysProValProSerSerLysArgSerSerGluLeuMetTyrAsnG 134
|||||
2899 CGCAAGCGGTGCATCATCAATTAGGAGCTCTGAACCTCATGTACAATG 2948

134 LuTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
2949 AGTACATCGTCTACACACATCCAGGTGAAGATGCAGTTCCTTGTCTGAAG 2998

151 ValArgPheHisLysLysArg 157
|||||
2999 GTGCGTTTCCATCAAGAGG 3019

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ60618

seq_documentation_block:
ID AAZ60618 standard; DNA; 3212 BP.

XX AC AAZ60618;

XX DT 16-MAY-2000 (first entry)

XX DE DNA encoding the poly(ADP-ribose) polymerase ZAP2 protein of Zea mays.

XX KW ZAP2; poly(ADP-ribose) polymerase; PARP; poly(ADP-ribose) transferase;
XX KW programmed cell death; apoptosis; growth rate; stress; cold; pathogen;
XX KW pest; drought; heat; fungi; nematode; seed-shatter; ss.

XX OS Zea mays.

XX FH Key Location/Qualifiers
XX CDS 81..3023

XX FT /*tag= a
XX FT /product= "ZAP2 protein"

XX PN WO200004173-A1.

XX XX 27-JAN-2000.

XX XX 12-JUL-1999; 99WO-EP04940.

XX XX 17-JUL-1998; 98US-0118276.

XX XX (PLBZ) PLANT GENETIC SYSTEMS NV.

XX PI Babychuk E, Kushnir S, De Block M;

XX XX WPI; 2000-182436/16.

XX DR P-PSDB; AAY68839.

XX XX Modulating cell death, growth and stress resistance in eukaryotes,
XX PT specifically plants, used, e.g. to impart fungus or nematode resistance

XX PT

XX XX

PS Disclosure; Page 103-108; 126pp; English.

XX

CC The present sequence encodes the ZAP2 protein of Zea mays. This protein
CC is a poly(ADP-ribose) polymerase (PARP) protein (also known as
CC poly(ADP-ribose) transferase). The PARP enzyme is involved in programmed
CC cell death or apoptosis, and is a nuclear enzyme. The ZAP2 polynucleotide
CC sequences can be used for modulation of programmed cell death in
CC eukaryotic cells. The method is used, specifically in plants, to induce,
CC or protect against, programmed cell death, depending on the extent to
CC which PARP activity is reduced. Reducing expression of endogenous ZAP
CC class PARP only is also used to modulate programmed cell death, to
CC increase growth rate and to produce plant cells that are more tolerant
CC of stress (cold, chemical treatments, pathogens, pests, drought, heat,
CC etc., or during transformation). Particular applications are generation
CC of plants that are resistant to fungi or nematodes; are male or female
CC sterile; or have better seed-shatter properties. The methods are also
CC used to improve growth of transformed plant cells (and derived calli or
CC complete plants).

SQ Sequence 3212 BP; 974 A; 600 C; 815 G; 823 T; 0 other;

alignment_scores:

Quality: 828.00 Length: 157
Ratio: 5.274 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:..

US-09-236-995D-4 x AAZ60618 ..

Align seg 1/1 to: AAZ60618 from: 1 to: 3212

```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17
|||||
2550 ACAACATGCTATTATGGCAGGTTCAAGTTGACGAATTTGTGGGAAT 2599

17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGly 34
|||||
2600 TCTTAGTCAAGGGCTAAGAAATTGCACCTCTCTCAGCACCTGTTACTGGCT 2649

34 yrmMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
2650 ATATGTCGCGCAAGGCTCTACTTTGCAGATCTAGTAAGCAAGAGCGCA 2699

51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
2700 CAATCTGTTATGTGGATAGGAATAATCCCTGATAGTTTGATGCTTCTTC 2749

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||
2750 TGAGTTGCTTTAGGAGACATGTATGAACATAACAAAGCCAGCTCCATGG 2799

84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
2800 ACAAACTCCCAAGAGGGAAGCATTCGACCAAGGATTTAGCAAAACCGTG 2849

101 ProLeuGluSerGluPheValLysTyrArgAspValValValProCy 117
|||||
2850 CCATGTGAGTCAGAGTTTGGAAGTGGAGGATGATGCTGTAGTTCCTG 2899

117 sGlyLysProValProSerSerLysArgSerSerGluLeuMetTyrAsnG 134
|||||
2900 CGCAAGCGGTGCATCATCAATTAGGAGCTCTGAACCTCATGTACAATG 2949

134 LuTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
2950 AGTACATCGTCTACACACATCCAGGTGAAGATGCAGTTCCTTGTCTGAAG 2999

151 ValArgPheHisLysLysArg 157
|||||
3000 GTGCGTTTCCATCAAGAGG 3020

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ23799

PN WO20007179-A2.
 XX 21-DEC-2000.
 XX 16-JUN-2000; 2000WO-US16629.
 XX 16-JUN-1999; 99US-0139543.
 XX (ICOS-) ICOS CORP.
 XX Christenson E, Demaggio AJ, Goldman PS, McElligott DL;
 XX WPI: 2001-025335/03.
 XX P-PSDB; AAB47031.
 XX New human poly(ADP-ribose) polymerase for treating inflammatory,
 PT neurological, cardiovascular, or neoplastic tissue growth disorders,
 PT such as, arthritis, encephalitis, myocardial ischemia, and leukocyte
 PT metastasis.
 XX
 XX Example 2; Page 101; 129pp; English.
 XX This sequence represents the 3' fragment of the human parp2
 CC (poly(ADP-ribose) polymerase) cDNA. This sequence was amplified
 CC using the primers given in AAC85315-17. DNA derived from a testis
 CC cDNA library was used as a template. The protein of the invention,
 CC hPARP2, causes the covalent addition of polymers of ADP-ribose to
 CC protein targets. hPARP2 activity is induced in many instances of
 CC oxidative stress or during inflammation where there is direct
 CC damage to the DNA. hPARP2 may be used to identify antagonists
 CC which may be used to treat a human having a disorder mediated by
 CC PARP2 activity, such as, inflammatory, neurological, cardiovascular,
 CC or neoplastic tissue growth disorders. hPARP2 and antibodies to it,
 CC can also be used to diagnose these conditions.
 XX
 XX Sequence 864 BP; 250 A; 207 C; 193 G; 214 T; 0 other;

alignment_scores:
 Quality: 441.50 Length: 158
 Ratio: 3.422 Gaps: 3
 Percent Similarity: 81.646 Percent Identity: 57.595
 alignment_block:
 US-09-236-995D-4 x AAC85320
 Align seg 1/1 to: AAC85320 from: 1 to: 864
 1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPhValGlyI 17
 376 AACAGGATGCTTCTATGTCATGCTGTTCCAGGATGAGTAAGTCCAGGAA 425
 17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
 426 CTGAGCCATGGGCTCGCAATTTGCCACCTGAAGCTCCATCCACAGTT 475
 34 yMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
 476 ACATGTTGGGAAGAAATCTACTTGTGTCATGCTCTTCCAGAGTGCC 525
 51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSe 67
 526 AATTACTGCTTTGCTCTCGCTAAAGAAATACAGGACTGCTCTTATC 575
 67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThr...SerM 83
 576 AGAGTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGT 625
 83 eTAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThr 99
 626 CCGAGGATTGCTTCAAGGTAACATAGCACCACAGGGCTGGCAGATG 675
 100 ValProLeuGluSerGluPheValLysTrpArgAspValValValPr 116

seq_name: /SIDS2/gcgdata/geneseq/genesep/NA2001.DAT:AAF59996
 seq_documentation_block:
 ID AAF59996 standard; cDNA; 1566 BP.
 XX
 AC AAF59996;
 XX
 DT 11-MAY-2001 (first entry)
 XX
 DE Human poly(ADP-ribose) synthetase sbhPARS2 cDNA.
 KW Human; poly(ADP-ribose) synthetase; sbhPARS2; vaccine; drug screening;
 KW ischaemic disorder; cerebral ischaemia; cardiac ischaemia;
 KW myocardial infarction; stroke; inflammation; autoimmune disease;
 KW diabetes; multiple sclerosis; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; chromosome localisation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key
 XX CDS
 FT 1..1566
 FT /*tag= a
 FT /product= "Human poly(ADP-ribose) synthetase sbhPARS2"
 XX
 WO200112645-A1.
 XX 22-FEB-2001.
 XX
 PF 10-AUG-2000; 2000WO-US21775.
 XX
 PR 12-AUG-1999; 99US-0373441.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Barone F, Field J, Kabnick KS, Li X, McQueney MS, Zhu Y;
 XX
 DR WPI: 2001-211196/21.
 DR P-PSDB; AAB60693.
 XX
 PT New human polypeptide of the polyADP-ribose synthetase family for use in
 PT diagnostic assays and for screening modulators used for preventing and
 PT treating inflammation, autoimmune disease and Alzheimers disease
 XX
 PS Claim 2; Page 21; 30pp; English.
 XX
 CC This sequence represents cDNA encoding a novel human poly(ADP-ribose)
 CC synthetase, sbhPARS2. The invention also relates to fragments,
 CC variants and sequences with at least 95% identity to the sbhPARS2
 CC protein or nucleotide sequence; expression systems and host cells
 CC comprising an sbhPARS2 nucleic acid sequence; the recombinant expression
 CC of sbhPARS2; and an antibody specific for sbhPARS2. sbhPARS2 proteins
 CC and nucleotides are useful as vaccines for inducing an immunological
 CC response in a mammal. The sbhPARS2 protein is useful for identifying
 CC compounds which inhibit or stimulate its activity or expression level.
 CC Such agonists and antagonists of sbhPARS2 are useful for treating human
 CC diseases including ischaemia and ischaemic tissue injury (e.g., cerebral
 CC and cardiac ischaemia, myocardial infarction, stroke), inflammation,
 CC autoimmune disease (e.g. diabetes, multiple sclerosis) and

CC neurodegenerative diseases (e.g., Parkinson's disease and Alzheimer's
 CC disease). shPARS2 nucleic acids are useful as diagnostic reagents for
 CC detecting mutations in the associated gene; as hybridisation probes
 CC to isolate full-length shPARS2 cDNAs and shPARS2 genomic clones; and
 CC for chromosome localisation studies. The shPARS2 protein is also useful
 CC as an immunogen to produce antibodies for therapeutic use. shPARS2
 CC proteins, nucleotides and antibodies are also useful in screening methods
 CC for detecting the effect of added compounds on the production of mRNA and
 CC protein in cells.

XX SQ Sequence 1566 BP; 490 A; 328 C; 381 G; 367 T; 0 other;

alignment_scores:
 Quality: 441.50 Length: 158
 Ratio: 3.422 Gaps: 3
 Percent Similarity: 81.646 Percent Identity: 57.595

alignment_block:
 US-09-236-995D-4 x AAF59996 ..

Align seg 1/1 to: AAF59996 from: 1 to: 1566

```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17
|||||
1078 AACAGGATGCTTCAAGGTAACAGTGGTCCAGGATGAGTAACGTGGGAAT 1127
|||||
17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||||
1128 CTGAGCCATGGGCTTCGAATGGCCACCTGAAGCTCCCATCACAGTT 1177
|||||
34 yrMetPheGlyGlyGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
1178 ACATGTTGGGAAGGAATCTACITGCTGACATGCTCTCCCAAGAGTGCC 1227
|||||
51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
1228 AATTACTGCTTGGCTCTCGCTTAAGAATAACAGGACTGCTCTTATC 1277
|||||
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysAlaThr...SerM 83
|||||
1278 AGAGTACTCTAGTCTAGTCTAATGACTACTAGAGGCCAATCTAAGG 1327
|||||
83 etAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThr 99
|||||
1328 CCGAAGGATGCTTCAAGGTAACATAGCACCAAGGGCTGGCAAGATG 1377
|||||
100 ValProLeuGluSerGluPheValLysTrpArgAspValValPr 116
|||||
1378 GCTCCAGTCTCTGCCACTGCTCACCCTGAATGGAGT...ACAGTGCC 1424
|||||
116 oCysGlyLysProValProSerSerIleArgSerSerGlu.....L 130
|||||
1425 AFTAGGACCAGCAGTGCACAGGAATCTGAATCCAGATGTTATACCC 1474
|||||
130 euMetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 146
|||||
1475 TCAACTACAATGAATATATTGATATACCCCAACCAAGTCCGATCGG 1524
|||||
147 PheLeuLeuLysValArgPheHis 154
|||||
1525 TACCTTTTAAAGGTTCAAGTTTAAAT 1548

```

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2001.DAT.AAC85303

seq_documentation_block:

ID AAC85303 standard; cDNA; 1814 BP.

XX AC AAC85303;

XX DT 29-MAR-2001 (first entry)

XX DE hparp2 cDNA.

XX Human; poly(ADP-ribose) polymerase; hPARP2; oxidative stress; ARDS;
 KW inflammation; ischaemic stroke; hemorrhagic shock; myocardial ischemia;
 KW infarction; cerebral vasospasm; rheumatoid arthritis; osteoarthritis;
 KW gouty arthritis; spondylitis; Behcet's disease; sepsis; septic shock;
 KW endotoxic shock; gram negative sepsis; gram positive sepsis; trauma;
 KW toxic shock syndrome; multiple organ injury syndrome; vasculitis;
 KW hemorrhage; conjunctivitis; uveitis; thyroid-associated ophthalmopathy;
 KW eosinophilic granuloma; asthma; chronic bronchitis; allergic rhinitis;
 KW chronic obstructive pulmonary disease; silicosis; reperfusion injury;
 KW pulmonary sarcoidosis; pleurisy; alveolitis; pneumonia; myocardium;
 KW bronchiectasis; pulmonary oxygen toxicity; keloid formation; brain;
 KW scar tissue formation; atherosclerosis; systemic lupus erythematosus;
 KW autoimmune thyroiditis; multiple sclerosis; Reynaud's syndrome;
 KW graft versus host disease; allograft rejection; cystic fibrosis;
 KW chronic glomerulonephritis; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; necrotizing enterocolitis;
 KW inflammatory dermatosis; contact dermatitis; atopic dermatitis;
 KW psoriasis; urticaria; fever; myalgia; meningitis; encephalitis;
 KW Sjogren's syndrome; alcoholic hepatitis; bacterial pneumonia;
 KW hypovolemic shock; Type 1 diabetes mellitus; hypersensitivity;
 KW leukocyte dyscrasia; thermal injury; cytokine-induced toxicity; ds.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 CDS 63..1814
 FT /*tag= a
 FT /product= "hPARP2"
 XX WO2000077179-A2.
 XX 21-DEC-2000.
 PD 16-JUN-2000; 2000WO-US16629.
 PE 16-JUN-1999; 99US-0139543.
 PR (ICOS-) ICOS CORP.
 XX Christenson E, Demaggio AJ, Goldman PS, McElligott DL;
 WPI; 2001-025335/03.
 DR P-PSDB; AAB47029.
 XX New human poly(ADP-ribose) polymerase for treating inflammatory,
 PT neurological, cardiovascular, or neoplastic tissue growth disorders,
 PT such as, arthritis, encephalitis, myocardial ischemia, and leukocyte
 PT metastasis
 XX Claim 1; Page 91-93; 129pp; English.
 PS This sequence encodes human poly(ADP-ribose) polymerase (hPARP2).
 CC This protein causes the covalent addition of polymers of ADP-ribose
 CC to protein targets. hPARP2 activity is induced in many instances of
 CC oxidative stress or during inflammation where there is direct damage
 CC to the DNA. hPARP2 may be used to identify antagonists which
 CC may be used to treat a human having a disorder mediated by PARP2
 CC activity, such as, inflammatory, neurological, cardiovascular,
 CC or neoplastic tissue growth disorders, e.g. ischemic stroke,
 CC hemorrhagic shock, myocardial ischemia or infarction,
 CC transplantation, cerebral vasospasm; rheumatoid, osteo- or gouty
 CC arthritis, spondylitis; Behcet's disease; sepsis, septic or endotoxic
 CC shock, gram negative or positive sepsis, toxic shock syndrome; multiple
 CC organ injury syndrome secondary to septicemia, trauma, or hemorrhage;
 CC allergic or vernal conjunctivitis, uveitis, thyroid-associated
 CC ophthalmopathy; eosinophilic granuloma; asthma, chronic bronchitis,
 CC allergic rhinitis, ARDS, chronic obstructive pulmonary disease,
 CC silicosis, pulmonary sarcoidosis, pleurisy, alveolitis, vasculitis,
 CC pneumonia, bronchiectasis, pulmonary oxygen toxicity; reperfusion
 CC injury of the myocardium, brain or extremities; cystic fibrosis; keloid
 CC formation, scar tissue formation; atherosclerosis; systemic lupus
 CC erythematosus, autoimmune thyroiditis, multiple sclerosis; Reynaud's

CC syndrome; graft versus host disease, allograft rejection; chronic
CC glomerulonephritis; inflammatory bowel disease, Crohn's disease,
CC ulcerative colitis; necrotizing enterocolitis; inflammatory dermatoses,
CC contact or atopic dermatitis, psoriasis, urticaria, fever and myalgias
CC due to infection; meningitis, encephalitis, and brain and spinal cord
CC injury due to minor trauma; Sjogren's syndrome; diseases involving
CC leukocyte diapedesis; alcoholic hepatitis; bacterial pneumonia;
CC antigen-antibody complex mediated diseases; hypovolemic shock; Type 1
CC diabetes mellitus; acute and delayed hypersensitivity; disease states
CC due to leukocyte dyscrasia and metastasis; thermal injury; granulocyte
CC transfusion associated syndromes; and cytokine-induced toxicity.
CC hPARP2 and antibodies to it, can also be used to diagnose these
CC conditions.
XX
SQ Sequence 1814 BP; 558 A; 389 C; 462 G; 405 T; 0 other;

alignment_scores:
Quality: 441.50 Length: 158
Ratio: 3.422 Gaps: 3
Percent Similarity: 81.646 Percent Identity: 57.595

alignment_block:

US-09-236-995D-4 x AAC85303 ..

Align seg 1/1 to: AAC85303 from: 1 to: 1814

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17
1326 AACAGGATGCTTCTATGGCATGGTTCAGGATGAGTAACCTGGTGGGAAT 1375
17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
1376 CTTGAGCCATGGCTTCGAATTCGCCACCTGAAGCTCCCATCACAGTT 1425
34 yMetPheGlyLeuGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
1426 ACATGTTTGGGAAGGAATCTACTTGTGTCGATGCTTCCCAAGAGTGCC 1475
51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeu 67
1476 AATTACTGCTTTCCTCTCGCTTAAGAAATACAGGACTGCTCTTATC 1525
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThr...SerM 83
1526 AGAGTAGCTCTAGTTCAGTGTGTAAGTAAGTACTAGAGGCCAATCCTAAG 1575
83 eAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThr 99
1576 CGGAAGGATTCCTCAAGGTAAACATACCAAGGGCTGGGCAAGATG 1625
100 ValProLeuGluSerGluPheValLysTrpArgAspValValValPr 116
1626 GCTCCAGTTCGCCCATCTGTCACCTGAATGGGAGT...ACAGTGCC 1672
116 oCysGlyLysProValProSerSerIleArgSerSerGlu.....L 130
1673 ATTAGGACCAAGCAAGTGCACAGGAATTCATCAATCCAGATGTTATACC 1722
130 eumetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 146
1723 TCAACTACATGAATATATTGTATATAACCAAGGCTGGTATCGG 1772
147 PheLeuLeuLysValArgPheHis 154
1773 TACCTTTTAAAGTTTCAGTTAAT 1796

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT.AA244287

seq_documentation_block:

ID_AA244287 standard; cDNA; 1843 BP.

XX

AC AA244287;

XX 31-MAR-2000 (first entry)
XX Human brain PARP2 cDNA.
DE
XX PARP; poly(ADP-ribose) polymerase; human; murine; detection; sepsis;
KW diagnosis; therapy; necrosis; apoptosis; neurodegenerative illness;
KW ischemic tissue damage; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3..1715
FT /*tag= a
FT /product= "PARP2"
XX
XX WO9964572-A2.
XX 16-DEC-1999.
XX
XX 04-JUN-1999; 99WO-EP03889.
XX
XX 05-JUN-1998; 98DE-1025213.
XX 01-MAR-1999; 99DE-1008837.
XX
XX (BADI) BASF AG.
XX
XX Kock M, Hoeger T, Kroeger B, Otterbach B, Lubisch W, Lemaire H;
XX WPI; 2000-087218/07.
XX P-PSDB; AAY51174.
XX
XX Novel genes and proteins, antibodies and binding partners useful in
XX diagnosis and therapy of energy deficiency associated disease
XX conditions
XX
XX Claim 7a; Page 49-52; 96pp; German.
XX
XX This invention describes novel human and murine poly(ADP-ribose)
XX polymerase (PARP) homologues, which are characterised by an amino acid
XX sequence with a functional NAD⁺-binding site and no zinc finger
XX sequence motif, of general formula CX₂CX₂MXH₂C₂ (1). The nucleic acid
XX sequences, PARP homologues and antibodies are useful for analytic
XX detection of PARP homologues and for identifying PARP effectors or
XX binding partners, as well as for determining their effectiveness.
XX PARP-binding partners are useful for the diagnosis or therapy of a
XX disease condition, which is the result of a PARP protein, especially an
XX energy deficiency, which may comprise tissue damage from cell death
XX following necrosis or apoptosis. The disease condition may be chosen
XX from a neurodegenerative illness, or sepsis or ischemic tissue damage,
XX in particular neurotoxic disturbances, etc. This sequence encodes the
XX human PARP2 protein used in the method of the invention.
XX
SQ Sequence 1843 BP; 599 A; 377 C; 447 G; 420 T; 0 other;

alignment_scores:

Quality: 441.50 Length: 158
Ratio: 3.422 Gaps: 3
Percent Similarity: 81.646 Percent Identity: 57.595

alignment_block:

US-09-236-995D-4 x AA244287 ..

Align seg 1/1 to: AA244287 from: 1 to: 1843

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17
1227 AACAGGATGCTTCTATGGCATGGTTCAGGATGAGTAACCTGGTGGGAAT 1276
17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
1277 CTTGAGCCATGGCTTCGAATTCGCCCATCTGAAGCTCCCATCACAGTT 1326

116 oCysGlyLysProValProSerSerIleArgSerSerGlu.....L 130
 1574 ATTAGGACCAGCAAGTGCACACAGGAATTCGAATCCAGATGGTTATACCC 1623
 130 eumetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 146
 1624 TCAACTACAATGAATATATATGCTATATACCCCAACCAAGGTCGCGATGCGG 1673
 147 PheLeuLeuLysValArgPheHis 154
 1674 TACCTTTAAAGGTCAGTTTAAT 1697

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77118

seq_documentation_block:

ID AAC77118 standard; cDNA; 1912 BP.

XX AAC77118;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2673 polynucleotide sequence SEQ ID NO:5345.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipariatic; antiparkinsonian; nontropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressive; cardiant;
 KW hypotensive; dermatological; coagulant; vasotropic; antidiabetic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antiinflammatory;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;

XX WPI; 2000-602362/57.

XX P-PSDB; ABA42909.

XX Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -

XX Claim 5; Page 4521-4522; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 which represent the human ORFX open reading frames 1 to 3161. The ORFX
 sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipariatic; antiparkinsonian; nontropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 1912 BP; 620 A; 390 C; 467 G; 433 T; 2 other;

alignment_scores:

Quality: 441.50 Length: 158

Ratio: 3.422 Gaps: 3

Percent Similarity: 81.646 Percent Identity: 57.595

alignment_block:

US-09-236-995D-4 x AAC77118 ..

Align seg 1/1 to: AAC77118 from: 1 to: 1912

1 AsnLysMetLeuLeuTyrHisGlySerArgLeuThrAsnPheValGlyI 17
 1291 AACAGGATGCTTCTATGTCATGGTTCAGGATGAGTAAGTGGTGGGAAT 1340

17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34

1341 CTTGAGCCATGGCTTCGAATTCGCCACCTGAAGCTCCCATCAGGTT 1390

34 yrMetPheGlyGlyGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50

1391 ACATGTTGGGAAAGGAATCTACTTTGCTGACATGCTCTCCACAGTGCC 1440

51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeu 67

1441 AATTACTGCTTTGGCTCTCGCCTAAAGAAATACAGGACTGCTGCTTATC 1490

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThr...SerM 83

1491 AGAGTAGCTCTAGTCAAGTGAATGAAGTACTAGAGGCCAATCCTAAGG 1540

83 etAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThr 99

1541 CCGAAGGATTGCTTCAAGGTAACATAGCACCAAGGGCTGGCAAGATG 1590

100 ValProLeuGluSerGluPheValLysTyrArgAspValValValPr 116

1591 GCTCCAGTCTGCCCCACTTCGCACCTCAATGGAGT...ACAGTGCC 1637

116 oCysGlyLysProValProSerSerIleArgSerSerGlu.....L 130

1638 ATTAGGACCAGCAAGTGCACACAGGAATTCGAATCCAGATGGTTATACCC 1687

130 eumetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 146

1688 TCAACTACAATGAATATATGCTATATACCCCAACCAAGGTCGCGATGCGG 1737

147 PheLeuLeuLysValArgPheHis 154

1738 TACCTTTAAAGGTCAGTTTAAT 1761

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC85341

seq_documentation_block:

ID AAC85341 standard; cDNA; 3200 BP.

XX AAC85341;

AC AAC85341;

OM of: US-09-236-995D-4 to: Issued_Patents_NA:* out_format : pfs
Date: Mar 9, 2002 7:46 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE1=framet p2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09236995/runat_08032002_113149_25118/app_query.fasta.1.1273
-DB=Issued_Patents_NA -OFMT=fastp -SUFFIX=rn1 -GAPOP=12.000
-GAPEX=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40 cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09236995_@cgn1_1_111 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLFPY
-WAIT -THREADS=1

Search information block:

Query: US-09-236-995D-4
Query length: 157
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 11323899
Search time (sec): 114.960000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len |
|--|-----------------|--------|---------|-------|
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-5 | 1 Documentation | 503.50 | 1160.85 | 3747 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-7 | | 435.50 | 993.59 | 5345 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-3 | | 429.50 | 987.30 | 2682 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-328-111-216 | | 358.50 | 834.70 | 595 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-4 | | 350.50 | 803.96 | 1592 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-196-387-1 | | 123.00 | 247.12 | 4134 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-196-387-7 | | 123.00 | 246.15 | 4491 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-196-387-9 | | 123.00 | 245.72 | 4657 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-522-229B-1 | | 77.00 | 146.72 | 1776 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-102-433-1 | | 75.00 | 146.72 | 1776 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-011-197-1 | | 75.00 | 139.78 | 2361 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-371-696-1 | | 72.50 | 135.91 | 1779 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-998-416-1056 | | 71.50 | 147.56 | 538 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-298-426-3 | | 69.50 | 128.74 | 1776 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-162-373-2 | | 69.00 | 118.21 | 3930 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-467-946-2 | | 69.00 | 118.21 | 3930 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-524-828-2 | | 68.50 | 127.62 | 1593 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-975-114A-2 | | 68.50 | 127.62 | 1593 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-849-281A-2 | | 68.50 | 127.62 | 1593 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-524-828-1 | | 68.50 | 123.57 | 2247 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-975-114A-1 | | 68.50 | 123.57 | 2247 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-658-136-2 | | 68.00 | 85.13 | 53526 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-658-136-1 | | 68.00 | 85.12 | 53577 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-620-467A-6 | | 67.00 | 136.11 | 569 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-348-572-6 | | 67.00 | 136.11 | 569 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-041-090B-6 | | 67.00 | 136.11 | 569 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-752-760A-1 | | 67.00 | 87.69 | 33081 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-427-497E-4 | | 66.00 | 115.86 | 3189 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-427-497E-3 | | 66.00 | 113.47 | 3189 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-341-843B-1 | | 66.00 | 111.49 | 3774 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-427-497E-1 | | 66.00 | 111.49 | 3774 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-427-497E-2 | | 66.00 | 111.49 | 3774 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-232-200-36 | | 65.50 | 113.44 | 2885 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-232-200-56 | | 65.50 | 113.44 | 2885 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-232-197-36 | | 65.50 | 113.44 | 2885 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-232-197-56 | | 65.50 | 113.44 | 2885 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-453-695A-111 | | 65.50 | 112.86 | 3033 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-268-161A-111 | | 65.50 | 112.86 | 3033 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-453-702A-111 | | 65.50 | 112.86 | 3033 |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-099-639-111 | | 65.50 | 112.86 | 3033 |

/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US95-08071-111 + 65.50 112.86 123.13
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-750-128-7 - 65.00 125.36 24.78
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-239-276-10 + 65.00 110.63 163.89
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-468-579B-10 + 65.00 110.63 163.89
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-468-577B-10 + 65.00 110.63 163.89
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-5
seq_documentation_block:
; Sequence 5, Application US/08044618
; Patent No. 5449605
; GENERAL INFORMATION:
; APPLICANT: SMULSON, MARK
; TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO
; TITLE OF INVENTION: CANCER BY THE USED OF RESTRICTION FRAGMENT LENGTH
; TITLE OF INVENTION: POLYMORPHISM OF THE GENE FOR THE HUMAN POLY (ADP-RIBOSE)
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/044,618
; FILING DATE: 19930406
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/257,696
; FILING DATE: 14-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: FOX, SAMUEL L
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0654.0490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)466-0800
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3747 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-044-618-5
alignment_scores:
Quality: 503.50 Length: 156
Ratio: 3.702 Gaps: 1
Percent Similarity: 87.179 Percent Identity: 60.897
alignment_block:
US-09-236-995D-4 x US-08-044-618-5 ..
Align seq 1/1 to: US-08-044-618-5 from: 1 to: 3747
1 AsnLysMetLeuThrHisGlySerArgLeuThrAsnPheValGlyI 17
2756 AACCCGAGATGCTGTGGCAGGGTCCAGGACCCACCACTTGTCTGGAT 2805
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
2806 CTTGTCCAGGGTCTCGGATAGCCGCGCTGAAGCCGCGTGCACAGCT 2855
34 yrMetPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerLysSerAla 50
|||||

[illegible]

seq_name: /cqn2 6/ptodata/2/ina/5A COMB.seq:US-08-044-618-7

```

seq_documentation_block:
; Sequence 7, Application US/08044618
; Patent No. 5449605
; GENERAL INFORMATION:
; APPLICANT: SWULSON, MARK
; TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO
; TITLE OF INVENTION: CANCER BY THE USED OF RESTRICTION FRAGMENT LENGTH
; TITLE OF INVENTION: POLYMORPHISM OF THE GENE FOR THE HUMAN POLY (ADP-RIBOSE)
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/044,618
; FILING DATE: 19930406
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/257,696
; FILING DATE: 14-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: FOX, SAMUEL L
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0654.0490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)466-0800
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5345 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA
US-08-044-618-7

alignment_scores:
  Quality: 435.50      Length: 156
  Ratio: 3.484        Gaps: 1
  Percent Similarity: 80.128  Percent Identity: 57.692

alignment_block:
US-09-236-995D-4 x US-08-044-618-7 ..

Align seg 1/1 to: US-08-044-618-7 from: 1 to: 5345

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17
|||
|||
3519 AACTGAAGGTTGCTGTGGCATGGTCCAGGACCCACCACTTTGCTGGAT 3568

17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||
|||
3569 CTTGTCCTCGGCTCTTTGGATGACCGCTGCTGAAGCACCTGTGATGGGT 3618

34 ytMetPheGlyLysGlyLeuTyPheAlaAspLeuValSerLysSerAla 50
|||
|||
3619 ACATGTGTTGGTAAGTGATCTATTTCGCTGATCTTCTCTCCAAAGATGCC 3668

51 GlnTyCysTyValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||
|||
3669 AACGACTGCCATACATCTTAGGAGACCCCAATAGGGTTAATCCTGTCGGA 3718

67 rgluValAlaLeuGlyAspMetTyrgluLeuLysLysAlaThrSerMet 84
|||
|||
3719 AGAAGTTGCCCTTGAAAGCTGTGTGAACCTGAAGCATGCTTCACATATCA 3768

84 sPLysProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||
|||
3769 GCAAGTTACCAAGGCAAGCACAGTGTCAAAGTTTGGGCACAACTACT 3818

101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||
|||
3819 CTTGACCTTTCAGTAGTATCCCA...CTGGAGTGGTGTAGAGTTCCTCT 3865

117 sGlyLysProValProSerSerIleArgSerGluLeuMetTyTyrAsn 134
|||
|||
3866 TGGGACCGGGTTTCATCTGGTGTGAATGACACCTCTCTACTGTATAATG 3915

134 luTyIleValTyRAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
|||
|||
3916 AGTACATTGCTATGATATGCTCAGGTAAATCTGAAATATCTGCTGAAA 3965

151 ValArgPheHisHisLys 156
|||
|||
3966 CTGAAATTCAAATTTTAAAG 3983

```

seq name: /cqn2 6/ptodata/2/ina/5A COMB.seq:US-08-044-618-3

```
seq_documentation_block:
: Sequence 3, Application US/08044618
```

```

/ Patent No. 5449605
/
/ GENERAL INFORMATION:
/
/ APPLICANT: SMULSON, MARK
/
/ TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO
/
/ TITLE OF INVENTION: CANCER BY THE USED OF RESTRICTION FRAGM
/
/ TITLE OF INVENTION: POLYMORPHISM OF THE GENE FOR THE HUMAN
/
/ TITLE OF INVENTION: POLYMERASE
/
/ NUMBER OF SEQUENCES: 7
/
/ CORRESPONDENCE ADDRESS:
/
/ ADDRESSEE: Sterne, Kessler, Goldstein & Fox
/
/ STREET: 1225 Connecticut Suite 300
/
/ CITY: Washington
/
/ STATE: D.C.
/
/ ZIP: 20036
/
/ COMPUTER READABLE FORM:
/

```



```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/044,618
; FILING DATE: 19930406
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/07/257,696
; FILING DATE: 14-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: FOX, SAMUEL L
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0654.0490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)466-0800
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2682 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-044-618-3

alignment_scores:
  Quality: 429.50      Length: 156
  Ratio: 3.492        Gaps: 1
Percent Similarity: 78.846 Percent Identity: 57.051

alignment_block:
US-09-236-995D-4 x US-08-044-618-3

Align seg 1/1 to: US-08-044-618-3 from: 1 to: 2682

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||
2050 AACTGAAGGTTGCTGTGGCATGGTCCAGGACCACCACTTTGCTGGGAT 2099

17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||||
2100 CCTGTCCTGGGTCTTGGATAGCCCTGCTGAAGCACCTGTGATGGGT 2149

34 yrMetPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerLysSerAla 50
|||||
2150 ACATGTTTGGTAAAGTGATCTATTTCGTGATCTTCTCCCAAGAGTGCC 2199

51 GlnTyrrCysTyrrValAspArgAsnAsnProValGlyLeuMetLeuSe 67
|||||
2200 ACGACTGCCATACATCTTAGGAAGACCAATAGGGTTAATCCTGTCGGA 2249

67 rGluValAlaLeuGlyAspMetTyrrGluLeuLysAlaThrSerMetA 84
|||||
2250 AGAAGTTGCCCTTGGAAAGCGTGTGTGAAGTGAAGCATGCTTCACATATCA 2299

84 sPlyProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
2300 GCAAGTTACCCAAAGGGCAAGCACAGTGTCAAAGTTTGGGCAAAACTACT 2349

101 ProLeuGluSerGluPheValLysTyrrPArgAspValValProCy 117
|||||
2350 CTGACCTTTTACGTAGTATCCCA...CTGGATGGGTAGAGGTTCCCTCT 2396

117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrrAsnG 134
|||||
2397 TGGGACACGGGTTTCTCTGTTGGTGTGAATCACACTGTCTACTTATATG 2446

134 luTyrrIleValTyrrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
2447 AGTACATTGTCTATGATATATGCTCAGGTAATCTGAAATATCTGCTGAAA 2496
```

```
151 ValArgPheHisHisLys 156
|||||
2497 CTCAAATTCATTTTAAAG 2514

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-328-111-216

seq_documentation_block:
; Sequence 216, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(595)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-216
```

```
alignment_scores:
  Quality: 358.50      Length: 108
  Ratio: 3.774        Gaps: 1
Percent Similarity: 87.963 Percent Identity: 64.815

alignment_block:
US-09-236-995D-4 x US-09-328-111-216/rev ..

Align seg 1/1 to reverse of: US-09-328-111-216 from: 1 to: 595

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||
342 AACAGGATGCTTCTATGGCATGGTTCAGGATGAGTAAGTGGTGGGAAT 293

17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||||
292 CTGTGACCATGGGCTTCGAATTGCCCATCTGAAGCTCCCATCACAGGTT 243

34 yrMetPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerLysSerAla 50
|||||
242 ACATGTTTGGGAAAGGAATCTACTTTGCTGACATGCTTCCCAAGAGTGCC 193

51 GlnTyrrCysTyrrValAspArgAsnAsnProValGlyLeuMetLeuSe 67
|||||
192 AATTACTGCTTTGCCCTCTCCCTAAAGAATACAGGACTGCTGCTCTATC 143

67 rGluValAlaLeuGlyAspMetTyrrGluLeuLysLysAlaThr...SerM 83
|||||
142 AGAGGTAGTCTAGGTTCAGTGTGAATGAAGTACTAGAGGCAATCCTTAAGG 93
```

83 eAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThr 99
::: :::
92 CGAAGGATTGCTCAAGTAACATAGACCAAGGGGCTGGCAAGATG 43

100 ValProLeuGluSerGluPheVal 107
::: :::::::::::::::::::::::
42 GTCGCCAGTTCTGCCACACTCGTC 19

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-4

seq_documentation_block:
; Sequence 4, Application US/08044618
; Patent No. 5449605
; GENERAL INFORMATION:
; APPLICANT: SMULSON, MARK
; TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO
; TITLE OF INVENTION: CANCER BY THE USE OF RESTRICTION FRAGMENT LENGTH
; TITLE OF INVENTION: POLYMORPHISM OF THE GENE FOR THE HUMAN POLY (ADP-RIBOSE)
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/044,618
; FILING DATE: 19930406
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/257,696
; FILING DATE: 14-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: FOX, SAMUEL L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0654, 0490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)466-0800
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1592 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-044-618-4

alignment_scores:
Quality: 350.50 Length: 128
Ratio: 3.505 Gaps: 1
Percent Similarity: 78.125 Percent Identity: 58.594

alignment_block:
US-09-236-995D-4 x US-08-044-618-4

Align seg 1/1 to: US-08-044-618-4 from: 1 to: 1592

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
1209 AACTGAAGGTGCTGCTGGCATGGTCCAGGACCACTTTGCTGGGAT 1258

17 eLeuSerGlnGlyLeuAlaProProGluAlaProValThrGlyT 34
1259 CCTGCTCGGGGCTCTTTGGATAGATCCCTGCCTGAAGCACCTGTGATGGGCT 1308

34 yMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
:::
1309 ACATGTTTGTAAAGTGATCTATTTTCGCTGATCTTCTCTCCAAGAGTGCC 1358

51 LcTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
::: :::
1359 AAGCACTGCCATACATCTTAGGAAGACCAATAGGTTAATCCTCTCGGA 1408

67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMet 84
:::
1409 AGAAGTTCCTTGGAAACGTGTGGAAGCATGCTTACATATCA 1458

84 sPlyProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
::: :::
1459 GCAAGTTACCAAGGCAAGCAGTGTCAAAGTTTGGCAAACTACT 1508

101 ProLeuGluSerGluPheValLysTrpArgAspValValProcy 117
::: :::
1509 CCTGACCTTTTCAGCTAGTAGTATCCA...CTGGATGGTGTAGAGGTTCTCT 1555

117 sGlyLysProValProSerSerIleArgSerSer 128
::: :::
1556 TGGGACCAAGGTTTCATCTGCTGTTGATGACACC 1589

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-387-1

seq_documentation_block:
; Sequence 1, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-09-196-387-1

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196.387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095.225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2027
; US-09-196-387-7

alignment_scores:
    Quality: 123.00      Length: 166
    Ratio: 1.577        Gaps: 7
    Percent Similarity: 46.988    Percent Identity: 25.904

alignment_block:
US-09-236-995D-4 x US-09-196-387-7 ..

Align seg 1/1 to: US-09-196-387-7 from: 1 to: 4491

1 AsnLysMetLeuLeuThrHisGlySerArgLeuThrAsnPheValGlyI 17
|||||::::::|||||:::|||||:::|||||:::|||||
3894 AATGAGCGCATGTTGTTTCATGGTTCTCTTTCATAAT.....GCCAT 3937

17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|::: :::|||||::: ::: :::
3938 TATTTCATAAAGGGTT.....GATCAGCGCATGCATACATAGGAG 3978

34 yrMetPheGlyLysGlyLeuTyrrPheAlaAspLeuValSerAla 50
|||||:::|||||:::|||||:::|||||:::|||||
3979 GAATGTTTGGGCGCGGATTTATTTTGCTGAAAACTCCTCAAAAAGCAAC 4028

51 GlnTyr..... 52
|||||

4029 CAATATCTTTATGGAAATGGAGGAGGAACAGCTGCCCTACACAAGGA 4078

53 .....CysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeu 67
|||||:::|||||:::|||||:::|||||:::|||||
4079 CAGTTCATGCTATATATGTCACACAA.....ATGCTCTTCT 4116

67 erGluValAlaLeuGlyAspMetTyrrGluLeuLysLysAlaThrSerMet 83
:::|||||:::|||||::: ::: :::
4117 GTAGAGTAGCCCTTGGGAAATCCTTCTGCGAGTTTAGCACCATGAAATG 4166

84 AspLysProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVa 100
||| ||| ||| ||| |||
4167 GCCACGCGCTCCAGGCGCCATCATGCTCAGTC..... 4196

100 lProLeuGluSerGluPheValLysTrpArgAspValValValProC 117

4197 ..... 4197

117 ysGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrrAsn 133
|||||:::||||| |||:::||||| ||| |||

```

```
4198 TTGGTAGACG.....AGCGTCATGG.....CTGGCATATGCT 4232
134 GluTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeu 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4233 GAATGATGTCATACAGAGGAGAACAGCATACCCAGAGTATCTATC 4280
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-387-9

seq_documentation_block:
; Sequence 9, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: Smith, Susan
; APPLICANT: de Lange, Titia
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIPI
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2855
US-09-196-387-9

alignment_scores:
Quality: 123.00 Length: 166
Ratio: 1.577 Gaps: 7
Percent Similarity: 46.988 Percent Identity: 25.904

alignment_block:
US-09-236-995d-4 x US-09-196-387-9 ..
Align seg 1/1 to: US-09-196-387-9 from: 1 to: 4657
1 AsnLysMetLeuLeuTyrPheHisGlySerArgLeuThrAsnPheValGlyI 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4060 AATGAGCGCATGTGTTTCATGGTCTCTCTTCAATTAAT.....GCCAT 4103
17 eLeuSerGlnGlyLeuAlaProProGluAlaProValThrGlyT 34
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4104 TATTATAAAGGGTTT.....CATGAGCGACATACATAGGAG 4144
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4145 GAATGTTTGGGCGCGGATTTATTTTTGGCTGAAAAAACCCTCAAAAACGAC 4194
51 GlnTyr..... 52
|||||
4195 CAATATCTTTTGAATTGGAGGAGAACAGGCTGCCCTACACACAAGGA 4244
53 .....CysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuS 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4245 CAGGTCATGCTATATATGTCACAGACAA.....ATGCTCTTCT 4282
67 erGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMet 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4283 GTAGAGTGACCCCTTGGAAATCTTCTTCTGAGTTTACACCACTGAAAATG 4332
84 AspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVa 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4333 GCCACGCGCTCCAGGGCACCAGTCAGTC..... 4362
100 lProLeuGluSerGluPheValLysTyrArgAspValValValProC 117
4363 .....A 4363
117 ysGlyLysProValProSerSerSerIleArgSerSerGluLeuMetTyrAsn 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4364 TTGTGAGACCG.....AGCGTCAATGGG.....CTGGCATATGCT 4398
134 GluTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeu 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4399 GAATATGTCATCTACAGAGGAGAACAGGATACCCAGAGTATCTTATC 4446
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-522-229B-1

seq_documentation_block:
; Sequence 1, Application US/08522229B
; Patent No. 5811291
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene Venke
; APPLICANT: Andersen, Lene No. 5811291boe
; APPLICANT: Dalboge, Henrik
; APPLICANT: Kauppinen, Markus Sakari
; APPLICANT: Christgau, Stephen
; APPLICANT: Heldt-Hansen, Hans Peter
; APPLICANT: Christoffersen, Claus
; APPLICANT: Nielsen, Per Munk
; APPLICANT: Voragen, Alphons Gerard Joseph
; TITLE OF INVENTION: An Enzyme With Rhamnogalacturonase Activity
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5811291o No. 5811291disk of No. 5811291th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/522,229B
; FILING DATE: 29-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 3953,204-US
; TELECOMMUNICATION INFORMATION:
```

APPLICANT: Andersen, Lene No. 5882911boe

```

48 sSer...AlaGlnTyrCysTyrValAspArgAsnAsnProVal..... 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
291 CACCTTGAGCGAGTACATGGTGTGCATTAATGGGACCGCAATCATTCACA 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 .GlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeu 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 TGGCGACATATATCACTCGCGAGCGGTCAATCGGCGAGCTGGCGTTCATC 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
78 LysLysAlaThrSerMetAspLysProProArgGlyLysHisSerThr... 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
391 GCTCGACTGAATTCGGACCTGTACCGAGGAGGCGGTTCGGCGACGTT 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
94 .....LysGlyLeuGlyLysThrValProLeuGluSerGluP 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 TCACCACCGCTGACGGAGCTG.....CCATTGAGGGATCAGA 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
106 heValLysTrpArgaspValValProCysGlyLysProValPro 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
479 TGTGTTTTTGG.....TCGGCAGTGAACCCCGCAGCA 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
123 SerSerIleArgSerSerGluLeu 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 AGTTCTACTAGAGCGAGCGATTTA 534
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-197-1

```

seq_documentation_block:
; Sequence 1, Application US/09011197
; Patent No. 6171789
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: No. 6171789el Insertion Sequence from a Virulent
; TITLE OF INVENTION: Isolate of Burkholderia Cepacia, and Diagnostic and
; TITLE OF INVENTION: Identification Procedures Based Thereon.
; NUMBER OF SEQUENCES: 16
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,398
; FILING DATE: 17-AUG-1995
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Burkholderia cepacia
; INDIVIDUAL ISOLATE: ET12/cb1a
US-09-011-197-1

```

```

alignment_scores:
  Quality: 75.50      Length: 180
  Ratio: 0.910       Gaps: 10
  Percent Similarity: 46.111  Percent Identity: 23.889

```

alignment_block:
US-09-236-995D-4 x US-09-011-197-1 ..

Align seg 1/1 to: US-09-011-197-1 from: 1 to: 2361

```

1 AsnLysMetLeuLeuTyrPHisGlySerArg.....Le 11
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1404 AACCTGCTGGCCAAAGTGGACCGGCTCTCGGCACGATTGGAAGAGCGGAT 1453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
11 uThrAsnPhe.....ValGlyIle..... 17
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1454 GACCCAGTTTCGGCTGCTTTACCCCGAACGATTCAACATTGGAATCTGAA 1503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
18 .....LeuSerGlnGlyLeu ArgIleAlaProProGlu..... 28
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1504 TCTCAACCGGCTCACACAGGAATCCGGATACCTCCACCCCAAGCGCGC 1553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29 .....AlaProValT 32
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1554 GCGCGAGAAAAACCCAGCGCGCTGCTGTTTCGAATCGCGCGCGCTGA 1603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
32 hrGlyTyrMetPhe.....GlyLysGlyLeuTyrPheAlaAspLeuVal 46
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1604 CCGCATCTCTGTTCTCAAGACCGGACTGGGCTGGCGGACCTGCC 1653
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
47 SerLys.....SerAlaGlnTyrCysTyr..... 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1654 GCCGAGATGGGCTGCGGCTCGGGCGTGACTTGCTGGCGGACGGCTACGCGA 1703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
55 .....ValAspArgAsnAsnProValGlyLeuMetL 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1704 TTGGCAGGCTGCGGGCGTATGGGATCGCCTGCACGAA.....TTGC 1744
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 euLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThr 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1745 TGCTTGGAAAGCTCGGAGCAGCAGACCAAAATCGATTTCTCGCGAGCGCA 1794
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
82 SerMetAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyL 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1795 GTCGATTCTCATCGATTCCGCGCGTGGGGCGGGCCCAAAACTGGGCCA 1844
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
98 sThrValProLeuGluSerGluPheValLysTrpArgAspAspValVal 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1845 AACCCCAACCGATCGG.....CGGACCAAGTCTTCAA 1876
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
115 alProCysGlyLysProValProSerSerIleArgSer 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1877 GCACCAATCGTCCGACGACGCAATGCGCGCGCTCG 1914
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-371-696-1

```

```

seq_documentation_block:
; Sequence 1, Application US/09371696
; Patent No. 6287777
; GENERAL INFORMATION:
; APPLICANT: Sytkowski, Arthur J.
; APPLICANT: Yang, Meiheng
; TITLE OF INVENTION: NOVEL NPG-1 GENE THAT IS DIFFERENTIALLY EXPRESSED IN PROSTATE
; TITLE OF INVENTION: TUMORS
; FILE REFERENCE: 01948/053002
; CURRENT APPLICATION NUMBER: US/09/371,696
; CURRENT FILING DATE: 1999-08-10
; EARLIER APPLICATION NUMBER: US 09/022,238
; EARLIER FILING DATE: 1998-02-11
; EARLIER APPLICATION NUMBER: US 08/644,326
; EARLIER FILING DATE: 1996-05-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1779
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (236)...(1225)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1779)

```


124 r.....lleargserSerGluLeuMetTyrAsnGluTyrI 136
442 CTAGGCGCGGCGCGCGCGGACCATTTAGTTCGACNCAATCNCATG 491
136 le 136
492 TC 493

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-298-426-3

seq_documentation_block:
; Sequence 3, Application US/08298426
; Patent No. 5516674
; GENERAL INFORMATION:
; APPLICANT: Roe, Richard M.
; APPLICANT: Hodgson, Ernest
; APPLICANT: Rose, Randy L.
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Degradation of Insecticides and Chlorinated Organics
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park &
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: NC
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,426
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-207A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Heliothis virescens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1680
; US-08-298-426-3

alignment_scores:
Quality: 69.50 Length: 202
Ratio: 0.808 Gaps: 11
Percent Similarity: 42.574 Percent Identity: 22.277

alignment_block:
US-09-236-995D-4 x US-08-298-426-3 ..
Align seg 1/1 to: US-08-298-426-3 from: 1 to: 1776
4 LeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyIleLeuSerGl 20
|||||
211 CTGATGGTGGTGGTTCATGACGAGCAGCAGTTCGTTGATACCCCTGGGGCG 260

20 nGlyLeuArgIleAlaProProGluAlaProValThrGlyTyrMetPheG 37
261 GTGTCTCAAGGCTTTCTCTGATGAAGATAGTAGGACACTAC..... 303
37 lyGlyLeuTyrPheAlaAspLeuValSer..... 47
304GACATGGTGGAGCCCTATCTTGGTAGTTCTC 333
48LysSerAlaGlnTyrCysTy 54
334 GACGTGGACACTGTCAAGAGGATCACCCTCAAGACTTTGAACAT...TT 380
54 rValAspArgAsn.....AsnProVal.....G 62
381 CGTTGACAGGGAACGTTCCACGAGCTTTGATCCCATCTTTGGAAAG 430
62 lyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLys 78
431 GGCTGCTCTGCTACAT.....GCTGACGAATGGAAGCAATG 468
79 LysAlaThr..... 81
469 CGGTCTAGATGAGTCCAGCGTTCACGAGCTCCAAAGATGCGCCTGATGGT 518
82SerMetAspLysProProArgGlyL 90
519 GCCCTTCATGGAGAGATCGCTTTGGAAATGATTAGTAGTACTCGGGGGA 568
90 ysHisSerThrLysGlyLeuGlyLys.....ThrValProLeuGlu 103
569 AG.....ATCAAGGATTCTGGGAAACCTTACATCGACGTGGAAGCCAAG 612
104 SerGluPheValLysTrpArgAspValValProCys.....G 118
613 AGTATGATGACCGAGTACGCGAATGACGTCATAGCCCTCATCGCCTTCGG 662
118 yLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnGluT 135
663 GTTGAAGTGAATCCCGAGCGGCGGACGAGCTTTATGTCAACAGTC 712
135 yrlleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLysVal 151
713 AAGCTATC.....ACCAAGTTTAAAGTTTTCAGCCTTCTCTGAAGTC 753
152 ArgPhe 153
754 CTGTTCT 759

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-162-373-2

seq_documentation_block:
; Sequence 2, Application US/09162373
; Patent No. 6013454
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/162,373
; CURRENT FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 1281811
US-09-162-373-2


```
alignment_scores:
  Quality: 69.00      Length: 70
  Ratio: 1.725      Gaps: 3
  Percent Similarity: 57.143      Percent Identity: 30.000

alignment_block:
US-09-236-995D-4 x US-09-162-373-2/rev ..

Align seg 1/1 to reverse of: US-09-162-373-2 from: 1 to: 3930

20 GlnGlyLeuArgIleAlaProProGluAlaProValThrGlyTyrMetPh 36
   |||||.....:|||||  |||  |||:
2399 CAGGGCCTCAATCTCAGCCGCCCGTGGCGGAGTCAGC..... 2361

36 eGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrC 53
   ::|||  |||||.....:  ::
2360 ....CAGGGCCACCCTCGTAGCAGATCTCCTCACCGC.....CTGCATC 2321

53 ysTyrValAspArgAsnAsnProValGlyLeuMetLeuSerGluVal 69
   |||:.....:|||||:|||||:  |||:  ::
2320 TTCAGGTCAGCCATGGTGGCCCGCAGCGGGGTCCTTGCCCTGCAGCCTGG 2271

70 AlaLeuGlyAspMetTyrGluLeuLysAlaThrSerMetAspLysPr 86
   |||||.....:  ::||
2270 CCGCTGGGGGATCTGATA.....AACCC 2248

86 oProArgGly 89
   |  |||||
2247 TCCGAGGGGC 2238
```



```

45 euValSerLysSerAlaGlnTyrCysTyrValAspArgAsnProVal 61
|||||
101 TAGTAAGCAAGAGTCACAGTATTCTTATGTGGATAGAAATAATCCGACT 150
|||||
62 GlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLeu 78
|||||
151 GCCTTGATGCTCTTCTGAGGTTCTCTAGGAGACATGCATGAAGTAA 200
|||||
78 slsAlaThrSerMetAspLysProProArgGlyLysHisSerThrLysG 95
|||||
201 AAAAGCAACGCCAATGGCAACACCTCCCAAGAGGAAGCATTCGACCAAG 250
|||||
95 LyLeuGlyLysThrValProLeuGluSerGluPheValLysTrpArgAsp 111
|||||
251 GCTTAGGCAAACTGTGCCACTAGAGTCGGAGTTTGTAAATGGAGGAT 300
|||||
112 AspValValProCysGlyLysProValProSerSerIleArgSerSe 128
|||||
301 GATGTCGTCGTCCTGTGGCAAGCCAGTCGACGATCATCAGGGCATC 350
|||||
128 rGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysM 145
|||||
351 TGAGCTTCTGTACAACGAGTATATAGTACAAACACAGCTCAGGTGAAGA 400
|||||
145 etGlnPheLeuLeuLysValArgPheHisHisLysArg 157
|||||
401 TGCAGTCTTGTGTGAAGGTCAAAATTCCTCACAAGCGT 438
|||||

seq_name: gb_est1:BE419432

seq_documentation_block:
LOCUS BE419432 536 bp mRNA EST 24-JUL-2000
DEFINITION WWS012.A12R000101 ITEC WWS Wheat Scutellum Library Triticum aestivum cDNA clone WWS012.A12, mRNA sequence.
ACCESSION BE419432
VERSION BE419432.1 GI:9417278
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticaceae; Triticum.
REFERENCE
1 (bases 1 to 536)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticaceae
JOURNAL Unpublished (2000)
COMMENT Contact: Schuch W
Zeneca Wheat Improvement Centre, Norwich Research Park
Colney Lane, Norwich NR4 7UH UNITED KINGDOM
Tel: 44 1603 250 2600
Fax: 44 1603 250 699
Email: wolfgang.schuch@aguk.zeneca.com
International Triticaceae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES
Location/Qualifiers
1..536
/organism="Triticum aestivum"
/cultivar="Novosibirskaya 67"
/db_xref="taxon:4565"
/clone="WWS012.A12"
/clone_lib="ITEC WWS Wheat Scutellum Library"
/tissue_type="scutellum callus"
/note="M13 Reverse sequencing primer used for 5' end of clone."
BASE COUNT 143 a 106 c 144 g 139 t 4 others
ORIGIN

```

```

alignment_scores:
Quality: 660.00 Length: 134
Ratio: 5.000 Gaps: 0
Percent Similarity: 98.507 Percent Identity: 91.791

alignment_block:
US-09-236-995D-4 x BE419432 ..
Align seg 1/1 to: BE419432 from: 1 to: 536

24 ileAlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLe 40
|||||
11 ATAGCACCTCCTCAGGCTCTGTGACAGGCTATATGTTGGCAAGGCCT 60
|||||
40 uTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAsp 57
|||||
61 CTACTTTGCAGATTAGTAAGCAAGAGTGCACAGTATTGTTATGTGATA 110
|||||
57 rGlnAsnProProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAsp 73
|||||
111 GAAATAATCCGACTGGCTTCTTCTTCTGAGGTTGCTCTAGGAGAC 160
|||||
74 MetTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLy 90
|||||
161 ATGCATGAACCTGAAAAGGCCAGCCCAATGACAAACCTCCAAGAGGAAA 210
|||||
90 sHisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheV 107
|||||
211 GCATTCGACCAAGGCTTAGGAAAACTGTGCCGCTAGAGTCGGAGTTTG 260
|||||
107 allYsTrpArqAspAspValValProCysGlyLysProValProSer 123
|||||
261 TGAATGGAGGATGATATCTGTCGCTTGTGGCAAGCCAGTCGCCAGCA 310
|||||
124 SerIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnTh 140
|||||
311 TCTATCANGGCATCTGAGCTTCTGTACACGAGTATATAGTCTACAACAC 360
|||||
140 rSerGlnValLysMetGlnPheLeuLeuLysValArgPheHisHisLys 157
|||||
361 AGCCCAAGGTGAAGATCGATCTTGTGTGAAGGTGAGATCCGTCACAAGC 410
|||||
157 rg 157
||
411 GT 412

seq_name: gb_est1:AV834168

seq_documentation_block:
LOCUS AV834168 661 bp mRNA EST 22-JUN-2001
DEFINITION AV834168 K. Sato unpublished cDNA library: Hordeum vulgare subsp. vulgare shoots germination Hordeum vulgare subsp. vulgare cDNA clone rbags5K04, mRNA sequence.
ACCESSION AV834168
VERSION AV834168.1 GI:14526257
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare.
ORGANISM Hordeum vulgare subsp. vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
REFERENCE
1 (bases 1 to 661)
AUTHORS Sato,K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2001)
COMMENT Contact: Kazuhiro Sato
Research Institute for Bioresources
Okayama University, Barley Germplasm Center
Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
Email: kzsato@rib.okayama-u.ac.jp,
URL:http://www.rib.okayama-u.ac.jp/barley/

```

Sato, K., Saisho, D., Takeda, K., Shini, T. and Kohara, Y. Direct submission;
 database: <http://www.shigen.nig.ac.jp/barley/Bartley.html>.

FEATURES

Location/Qualifiers
 1..661
 /organism="Hordeum vulgare subsp. vulgare"
 /cultivar="Haruna Nijo"
 /db_xref="taxon:112509"
 /clone_lib="K. Sato unpublished cDNA library: Hordeum vulgare subsp. vulgare shoots germination"
 /tissue_type="shoots"
 /dev_stage="germination"
 180 a 181 c 132 g 168 t

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 642.00 Length: 136
 Ratio: 4.827 Gaps: 1
 Percent Similarity: 97.794 Percent Identity: 90.441

alignment_block:

US-09-236-995D-4 x AV834168/rev ..

Align seg 1/1 to reverse of: AV834168 from: 1 to: 661

23 ArgileAlaProGluAlaProValThr.GlyTyrMetPheGlyLysG 39
 |||||
 661 AGGATAGCACCCTCTGAGGACCCGCTGACCGGGCTATATGTTGGCAAAG 612
 |||||
 39 lYLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrVal 55
 |||||
 611 GCCTCTACTTTGACAGATTAGTAGCAAGAGTGCACAGTATTGTTATG 562
 |||||
 56 AspArgAsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuG 72
 |||||
 561 GATAGAAAAATCCGACTGGCTGATGCTCTTCTGAGGTTGCTCTAGG 512
 |||||
 72 yAspMetTyrGluLeuLysLysAlaThrSerMetAspLysProProArg 89
 |||||
 511 AGACATTTCATGAACCTGAAAAAGAACGACCCCAATGGACAAACCTCCAAG 462
 |||||
 89 lYlYsHisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGlu 105
 |||||
 461 GGAAGCATTCGACCAAGGCTTAGGCAAACTGTGCCACTAGAGTCGAG 412
 |||||
 106 PheValLysTrpArgAspValValValProCysGlyLysProValPr 122
 |||||
 411 TTTCTTAATGGAGGATGATGCTGCTGCTTGTGGCAAGCCAGTGCC 362
 |||||
 122 oSerSerIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrA 139
 |||||
 361 AGCATCTATCAGGCTCTGAGCTTCTGTACACGAGTATATAGTGATACA 312
 |||||
 139 snThrSerGlnValLysMetGlnPheLeuLeuLysValArgPheHisHis 155
 |||||
 311 ACACAGCTCAGTGAAGATGAGTCTTCTTGTGAAGTCAAAATCCGTCAC 262
 |||||
 156 LysArg 157
 |||||
 261 AAGCGT 256

seq_name: gb_estl:BE420229

seq_documentation_block:

LOCUS BE420229 369 bp mRNA EST 24-JUL-2000
 DEFINITION WWS04.C11R000101 ITEC WWS Wheat Scutellum Library Triticum aestivum
 CDNA clone WWS04.C11, mRNA sequence.
 ACCESSION BE420229
 VERSION BE420229.1 GI:9418075
 KEYWORDS EST.
 SOURCE bread wheat.

ORGANISM

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Triticum.

REFERENCE

AUTHORS

1 (bases 1 to 369)
 Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,
 S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
 Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
 Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
 Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
 Sorrells, M., Warburton, M. and Wenzel, G.

TITLE

International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae

JOURNAL

COMMENT

Contact: Schuch W
 Unpublished (2000)
 Zeneca Wheat Improvement Centre, Norwich Research Park
 Colney Lane, Norwich NR4 7UH UNITED KINGDOM
 Tel: 44 1603 250 2600
 Fax: 44 1603 250 699
 Email: wolfgang.schuch@aguk.zeneca.com
 International Triticeae EST Cooperative (ITEC)
<http://wheat.pw.usda.gov/genome>.

FEATURES

source

1..369
 /organism="Triticum aestivum"
 /cultivar="Novosibirskaya 67"
 /db_xref="taxon:4565"
 /clone="WWS04.C11"
 /clone_lib="ITEC WWS Wheat Scutellum Library"
 /tissue_type="scutellum callus"
 /note="M13 Reverse sequencing primer used for 5' end of
 clone."
 103 a 78 c 101 g 87 t

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 569.00 Length: 120
 Ratio: 4.822 Gaps: 0
 Percent Similarity: 98.333 Percent Identity: 91.667

alignment_block:

US-09-236-995D-4 x BE420229 ..

Align seg 1/1 to: BE420229 from: 1 to: 369

24 IleAlaProGluAlaProValThrGlyTyrMetPheGlyLysGlyLe 40
 |||||
 11 ATAGCACCTCCCTGAGGCTCCTGTGACAGCTATATGTTGGCAAGGCT 60
 |||||
 40 uTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAsp 57
 |||||
 61 CTACTTGCAGATTAGTAGAAGCAAGAGTGCACAGTATTGTTATGTTG 110
 |||||
 57 rgAsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuGly 73
 |||||
 111 GAAAAATCCGACTGGCTGATGCTCTTCTGAGGTTGCTCTAGGAGAC 160
 |||||
 74 MetTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGly 90
 |||||
 161 ATGCATGAACCTGAAAAAGGCCACCGCAATGGACAAACCTCCCAAGAG 210
 |||||
 90 sHisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPhe 107
 |||||
 211 GCATTCCACCAAGGCTTAGGAAAACTGTCCGCTAGAGTCGGAGTTG 260
 |||||
 107 alLysTrpArgAspValValValProCysGlyLysProValProser 123
 |||||
 261 TGAATGGAGGATGATATCGTCGCTTGTGGCAAGCCAGTGCAGCA 310
 |||||
 124 SerIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsn 140
 |||||
 311 TCTATCAGGGCATCTGAGCTTCTGTC.AACGAGTATATAGTACAAAC 359

alignment_scores:

Quality: 517.50 Length: 156
Ratio: 3.777 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:

US-09-236-995D-4 x BE304535 ..

Align seg 1/1 to: BE304535 from: 1 to: 571

```

1 AsnLysMetLeuLeuTyrPheHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 AACCCGAAGATTGCTGTGGCAGCGGTCCAGGACCACCAACTTTGCTGGGAT 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 CCTGCCAGGGTCTCGGATAGCCGCCCTGAAGCCCGCTGACAGGCT 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 ACATGTTGGTAAAGGATCTATTTCGCTGACATGGCTCCACAGAGTGC 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuSe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 AACTACTGCCATACGCTCAGGAGAGCCCAATAGGCTTAATCCTGTTGG 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysAlaThrSerMetA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 AGAAGTTGCCCTTGGAAACATGTATGAACCTGAACGCGCTTCACATATCA 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 GCAAGTTACCCAGGCAAGCAGCAGTGTCAAAAGGTTTGGGCAAACTACC 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 CCTGATCCTTCAGTACATAGT...CTGGATGGTGTAGACGTTCTCT 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 TGGACCGGGATTTCATCTGTGTGTGAATGACACCTCTCTACTATATAAG 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 LuTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
437 AGTACATTGTCTATGATATTGCTCAGGTAAATCTGAAGTATCTGCTGAAA 486
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 ValArgPheHisLysLys 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
487 CTGAAATTCAAATTTTAAG 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: gb_estl:AA397988

seq_documentation_block:

LOCUS AA397988 579 bp mRNA EST 12-AUG-1997
DEFINITION zt86c12.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:729238
5' similar to gb:J03473 NAD(+) ADP-RIBOSYLTRANSFERASE (HUMAN);,
mRNA sequence.
ACCESSION AA397988
VERSION AA397988.1 GI:2050655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 579)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubouque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,
T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
CONTACT: Wilson RK
COMMENT Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1320 Std Error: 0.00

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 427.

FEATURES

source

1..579
/organism="Homo sapiens"
/db_xref="GDB:5925767"
/db_xref="taxon:9606"
/clone="IMAGE:729238"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc. and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 148 a 138 c 144 g 149 t
ORIGIN

alignment_scores:

Quality: 517.50 Length: 156
Ratio: 3.777 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:

US-09-236-995D-4 x AA397988 ..

Align seg 1/1 to: AA397988 from: 1 to: 579

```

1 AsnLysMetLeuLeuTyrPheHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 AACCCGAAGATTGCTGTGGCAGCGGTCCAGGACCACCAACTTTGCTGGGAT 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 CCTGTCACAGGGTCTTCGGATAGCCCGCTGAAGCGCCCGTGACAGGCT 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 ACATGTTGGTAAAGGATCTATTTCGCTGACATGGCTCCACAGAGTGC 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 GlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuSe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 AACTACTGCCATACGCTCAGGAGAGCCCAATAGGCTTAATCCTGTTGG 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 AGAAGTTGCCCTTGGAAACATGTATGAACGACGCTTCACATATCA 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 sPLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 GCAAGTTACCCAGGCAAGCAGCAGTGTCAAGGTTTGGGCAAACTACC 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 CCTGATCCTTCAGCTAACATAGT...CTGGATGGTGTAGACGTTCTCT 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 TGGGACCGGGATTTCATCTGGTGTGAATGACACCTCTCTACTATATAAG 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

134 luTyrIleValtyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
|||||
440 AGTACATTGCTATGATATTCCTCAGGTAAATCTGAAGTATCTGCTCAA 489

151 ValArgPheHisLysLys 156
|||||
490 CTGAAATCAATTTTAAG 507

seq_name: gb_est1:AA401836

seq_documentation_block: 669 bp mRNA EST 16-MAY-1997
LOCUS AA401836
DEFINITION zV66b12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:758591 5' similar to gb:J03473 NAD(+) ADP-RIBOSYLTRANSFERASE
(HUMAN); mRNA sequence.

ACCESSION AA401836
VERSION AA401836.1 GI:2055855
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 669)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,
T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997

JOURNAL Unpublished (1997)
COMMENT Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 Et from Amersham
High quality sequence stop: 493.

FEATURES

Location/Qualifiers
1..669
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:758591"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCGAATGGAGCGCGCTTAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 191 a 152 c 157 g 169 t
ORIGIN

alignment_scores:
Quality: 517.50 Length: 156
Ratio: 3.777 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:

US-09-236-995D-4 x AA401836 ..

Align seg 1/1 to: AA401836 from: 1 to: 669

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 17

|||||
200 AACCGAAGATTGCTGGCACGGGTCCAGGACCACCACTTTGCTGGGAT 249
17 eLeuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGlyT 34
|||||
250 CCTGTCCCGAGGGCTTCGGATAGCCCGCCGCTGAAGCCCGCTGACAGGT 299
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
300 ACATGTTTGGTAAAGGATCTATTTGCTGACATGGTCTCCAAGAGTGCC 349
51 GlnTyrCystTyrValAspArgAsnProValGlyLeuMetLeuLeu 67
|||||
350 AACTACTGCATACGCTCTCAGGAGACCAATAGGCTTAATCCTGTTGGG 399
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMet 84
|||||
400 AGAAGTTGCCCTTGGAAACATGATGAACCTGAAGCAGCGTTCACATATCA 449
84 sPlysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
450 GCAAGTTACCCCAAGGCGCAAGCAGGTGTCAAAGTTTGGGCAAACTACC 499
101 ProLeuGluSerGluPheValLysTrpArgAspValValProCy 117
|||||
500 CCTGATCCTCTCAGCTAACATTAGT...CTGGATGGTGTAGAGCTTCCTCT 546
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||
547 TGGGACCGGGATTTCATCTGGTGTGAATGACACCTCTCTACTATATAACG 596
134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
|||||
597 AGTACATTGCTATGATATTCCTCAGGTAAATCTGAAGTATCTGCTGAAA 646
151 ValArgPheHisLysLys 156
|||||
647 CTGAAATCAATTTTAAG 664

seq_name: gb_est2:BI093436

seq_documentation_block:
LOCUS BI093436 916 bp mRNA EST 20-JUN-2001
DEFINITION 602859629F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5000911 5',
mRNA sequence.

ACCESSION BI093436
VERSION BI093436.1 GI:14511766
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 916)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM1033 row: i column: 08

High quality sequence stop: 837.

FEATURES

source

1..916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5000911"
/clone_lib="NIH_MGC_10"

BASE COUNT 248 a 209 c 230 g 229 t
ORIGIN

/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

alignment_scores:
Quality: 517.50 Length: 156
Ratio: 3.777 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:
US-09-236-995D-4 x BI093436 ..
Align seg 1/1 to: BI093436 from: 1 to: 916

```

1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||
185 AACCGAAGATTCCTGGCAGCGGTCCAGGACCACTTTCCTGGGAT 234

17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||
235 CCTGTCCAGGGTCTTCGGATAGCCCGCTGAAGCGCGGTGACAGGCT 284

34 yMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
285 ACATGTTTGTAAAGGATCTATTTTCCTGACATGGTCCCAAGAGTGC 334

51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
335 AACTACTGCCATACGCTCAGGAGACCAATAGGCTTAATCTGTGGG 384

67 xGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||
385 AGAAGTTGCCCTTGGAAACATGTATGAACCTGAAGCAGCTTCACATCA 434

84 sPlysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
435 GCAAGTTACCAAGGCAACACAGTGTCAAGGTTTGGGCAAACTACC 484

101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||||
485 CCTGATCCTTCAGCTAACATTAGT...CTGGATGGGTAGACGTTCTCT 531

117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||
532 TGGGACCGGATTCATCTCGTGTGAATGACACCTCTCTACTATAAAG 581

134 luTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
582 AGTACATTGTATGATATGCTCAGGTAATCTGAAGTATCTGCTGAAA 631

151 ValArgPheHisLys 156
|||||
632 CTGAATTCATTAAG 649

```

seq_name: gb_est2:BF793705

seq_documentation_block:
LOCUS BF793705 952 bp mRNA EST 12-JAN-2001
DEFINITION 602255121F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4347365 5',
mRNA sequence.
ACCESSION BF793705
VERSION BF793705.1 GI:12098759
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 952)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM9970 row: j column: 06
High quality sequence stop: 672.

FEATURES

source
1..952
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4347365"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; Oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 231 a 215 c 244 g 262 t
ORIGIN

alignment_scores:

Quality: 517.50 Length: 156
Ratio: 3.777 Gaps: 1
Percent Similarity: 87.821 Percent Identity: 62.179

alignment_block:

US-09-236-995D-4 x BF793705 ..
Align seg 1/1 to: BF793705 from: 1 to: 952
1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI1 17
|||||
52 AACCGAAGATTCCTGGCAGCGGTCCAGGACCACTTTCCTGGGAT 101
17 eLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValThrGlyT 34
|||||
102 CCTGTCCAGGGTCTTCGGATAGCCCGCTGAAGCGCGGTGACAGGCT 151
34 yMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
152 ACATGTTTGTAAAGGATCTATTTTCCTGACATGGTCTCCAAGAGTGC 201
51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
202 AACTACTGCCATACGCTCAGGAGACCACTTTCCTGTGGG 251
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
|||||
252 AGAAGTTGCCCTTGGAAACATGTATGAACCTGAAGCAGCTTCACATCA 301
84 sPlysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
302 GCAAGTTACCAAGGCAACACAGTGTCAAGGTTTGGGCAAACTACC 351
101 ProLeuGluSerGluPheValLysTrpArgAspValValValProCy 117
|||||
352 CCTGATCCTTCAGCTAACATTAGT...CTGGATGGGTAGACGTTCTCT 398
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||
399 TGGGACCGGGATTCATCTGGTGAATGACACCTCTCTACTATAAAG 448

```
134 luTyrileValTyrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
449 AGTACATTGCTATGATATGCTCAGGTAATCTGAAGTATCTGCTGAAA 498
151 ValArgPheHisLys 156
499 CTGAATTCATTTTAAAG 516
seq_name: gb_est2:BE911127
```

```
seq_documentation_block:
LOCUS BE911127 587 bp mRNA EST 29-SEP-2000
DEFINITION 601662539F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962696 5',
mRNA sequence.
ACCESSION BE911127
VERSION BE911127.1 GI:10408111
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 587)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9129 row: n column: 09
High quality sequence stop: 559.
Location/Qualifiers
1..587
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3962696"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_host="3 months, virgin"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
```

```
BASE COUNT 149 a 163 c 150 g 125 t
ORIGIN
```

```
alignment_scores:
Quality: 513.50 Length: 156
Ratio: 3.776 Gaps: 1
Percent Similarity: 87.179 Percent Identity: 61.538
```

```
alignment_block:
US-09-236-995D-4 x BE911127 ..
Align seg 1/1 to: BE911127 from: 1 to: 587
```

```
1 AsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyTl 17
|||||
25 AACCGGAGGCTGTGTGGCAGCGCTCCAGGACCACTTTGCTGGCAT 74
17 eLeuSerGlnGlyLeuArgLysAlaProGluAlaProValThrGlyT 34
|||||
75 CCATGCGCAGGCTGTGGGTAGTACCCCACTGAGCGCTGTGACAGGT 124
```

```
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
|||||
125 ACATGTTGGAAAGGATCTACTTTGCCGACATGTTGCCAAAGTGA 174
51 GlnTyrCysTyrValAspArgAsnProValGlyLeuMetLeuLeuSe 67
|||||
175 AACTACTGCCACACATCTCAGGAGACCCGATTGGCTTAATATCTGCTGG 224
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMet 84
|||||
225 AGAGTTGCCCTTGGAAACATGATGAACCTCAAGCATGCTTCACATATCA 274
84 sPlyProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
|||||
275 GCAAGTTACCCCAAGGCAAGCAGTGTCAAAGTTTGGGAAAACACACC 324
101 ProLeuGluSerGluPheValLysTyrArgAspValValProCy 117
|||||
325 CCTGACCTTTCGGCCACATCACC...CTGGAGGGGTGTAGAGTTCCACT 371
117 sGlyLysProValProSerSerIleArgSerSerGluLeuMetTyrAsnG 134
|||||
372 GGAACAGGATCCCATCTGCTGTCACGACACCTGCTGCTGTATCATG 421
134 luTyrileValTyrAsnThrSerGlnValLysMetGlnPheLeuLys 150
|||||
422 AGTACATTGCTACGACATGCTCAGGTGAATCTCAAAATACCTGCTGAAA 471
151 ValArgPheHisLys 156
|||||
472 CTCAAGTTCAATTTTAAAG 489
seq_name: gb_est1:BE382739
```

```
seq_documentation_block:
LOCUS BE382739 709 bp mRNA EST 21-JUL-2000
DEFINITION 601297572F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3627881 5',
mRNA sequence.
ACCESSION BE382739
VERSION BE382739.1 GI:9328104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 709)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM311 row: o column: 18
High quality sequence stop: 646.
Location/Qualifiers
1..709
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3627881"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
```

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library."

BASE COUNT 178 a 167 c 181 g 183 t
ORIGIN

alignment_scores:
Quality: 510.50 Length: 156
Ratio: 3.754 Gaps: 1
Percent Similarity: 87.179 Percent Identity: 61.538

alignment_block:
US-09-236-995D-4 x BE382739 ..

Align seg 1/1 to: BE382739 from: 1 to: 709

```
1 AsnLysMetLeuLeuTyrHisGlySerArgLeuThrAsnPheValGlyI1 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
12 AACCGAAGATTGCTGGCAGCGGTCCAGGACCAACCTTGTCTGGGAT 61
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 eLeuSerGlnGlyLeuArgLeuAlaProProGluAlaProValThrGlyT 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
62 CCTGTCAGGATCTCGGATAGCCCGCTGAAGCCCGCTGACAGGCT 111
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 yrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAla 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
112 ACATGTTTGGTAAAGGATCTATTTCGCTGACATGGTCTCCAAGAGTGCC 161
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GlnTyrCysTyrValAspArgAsnAspProValGlyLeuMetLeuLeuSe 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
162 AACTACTGCGCATAGCTCTAGGAGAGCCCAATAGGCTTAATCTGTGGG 211
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 rGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetA 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
212 AGAAGTTGCCCTTGGAAACATGATGAACCTGAAGCAGCTTCACATATCA 261
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 sPlysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVal 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
262 GCAAGTTACCAACCAAGGCAAGCAGTGTCAAGGTTTGGGCAAACTACC 311
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ProLeuGluSerGluPheValLysTyrArgAspAspValValProCy 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
312 CTGATCTTCACTACATTAAGT...CTGGATGGTGTAGACCTTCTCT 358
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 sGlyLysProValProSerSerLysArgSerGluLeuMetTyrAsnG 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
359 TGGGACCGGATTTTCATCTGCTGTGAATGACACCTCTCTACTATATA 408
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 lutyryleValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLys 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
409 AGTACATTGCTATGATATTGCTCAGGTAATCTGAAGTATCTGCTGAAA 458
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ValArgPheHisLys 156
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
459 CTGAATTCATTTTAAAG 476
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

seq_name: gb_est2:BI079902

seq_documentation_block:
LOCUS BI079902 725 bp mRNA EST 20-JUN-2001
DEFINITION 602876096F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5007983 5',
mRNA sequence.
ACCESSION BI079902
VERSION BI079902.1 GI:14498232
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 725)
NH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11051 row: o column: 24
High quality sequence stop: 664.

FEATURES

source
1..725
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5007983"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 177 a 204 c 188 g 156 t
ORIGIN

alignment_scores:

Quality: 509.50 Length: 153
Ratio: 3.802 Gaps: 1
Percent Similarity: 87.582 Percent Identity: 62.745

alignment_block:

US-09-236-995D-4 x BI079902 ..
Align seg 1/1 to: BI079902 from: 1 to: 725

```
4 LeuLeuTyrHisGlySerArgLeuThrAsnPheValGlyLysLeuSerG1 20
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3 CTGCTGTGGCAGCGCTCCAGGACCAACCTTGTCTGGCCTCTGTCGCA 52
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
20 nGlyLeuArgLeuAlaProProGluAlaProValThrGlyTyrMetPheG 37
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
53 GGCTCTGGGATAGCCCACTGAAGCCCTGTGACAGCTACATGTTTG 102
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
37 lYlysGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCys 53
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
103 GGAAGGGATCTACTTTCGACATGGTGTCCAAAGTGCAAACTACTGC 152
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
54 TyrValAspArgAsnAspProValGlyLeuMetLeuLeuSerGluValAl 70
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
153 CACACATCTCAGGGAGACCGGATTTGGCTTAATCTGCTGGGAGAGTTGC 202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
70 aLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetAspLysProp 87
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
203 CCTTGGAAACATGATGAACCTCAAGCATGCTTCACATATCAGCAAGTTAC 252
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
87 roArgGlyLysHisSerThrLysGlyLeuGlyLysThrValProLeuGlu 103
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
253 CCAAGGGCAAGCACAGTGTCAAAGGTTTGGGAAAACCAACCCCTGACCC 302
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
104 SerGluPheValLysTyrArgAspValValValProCysGlyLysPr 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
303 TCGGCCAGCATCACC...CTGGAGGGTGTAGAGGTTCTCCACTGGGAACAGG 349
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
120 oValProSerSerLysArgSerGluLeuMetTyrAsnGluTyrIlev 137
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
350 GATCCCATCTGGTGTACAGCACCTGCTGCTGCTATATAGTACATG 399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
137 alTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLysValArgPhe 153
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

400 TCTACGACATTGCTCAGTGGAATCTCAAAATACCTGCTGCAAACTCAAGTTC 449

154 HisHisLys 156
:::
450 AATCTTAAG 458

seq_name: gb_est2:BG675913

seq_documentation_block: 928 bp mRNA EST 01-MAY-2001
LOCUS BG675913 6026226263f1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4747574 5',
DEFINITION mRNA sequence.
ACCESSION BG675913
VERSION BG675913.1 GI:13907309
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 928)
NTI-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: Cgaps-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10597 row: m column: 15
High quality sequence stop: 816.
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4747574"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 247 a 209 c 225 g 247 t
ORIGIN
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4747574"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source

alignment_scores:
Quality: 503.50 Length: 153
Ratio: 3.757 Gaps: 1
Percent Similarity: 87.582 Percent Identity: 62.092
alignment_block:
US-09-236-995D-4 x BG675913
Align seg 1/1 to: BG675913 from: 1 to: 928

4 LeuLeuTrpHisGlySerArgLeuThrAsnPhleValGlyLeuSerG1 20
6 TTGCTGTGGCAGCGGTCCAGGACCACTTTGCTGGGATCTCTGTCCTCA 55
20 nGlyLeuArgIleAlaProGluAlaProValThrGlyTyrMetPheG 37
56 GGATCTTCGGATAGCCCGCTGAGCGCCGCTGAGCGGATCTCTGTCCTCA 105
37 lYlsGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCys 53
106 GTRAAAGGATCTATTTCGCTGACATGCTCCCAAGAGTGCCAACTACTGC 155
54 TyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSerGluValAl 70

156 CATACGCTCTCAGGAGACCAATAGCTTAATCCTCTGTTGGAGAAAGTTC 205
70 aLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetAspLysProp 87
206 CCTTGGAAACATCATGACCTGAAGCAGCTTCACATATCAGCAAGTTC 255
87 roArgGlyLysHisSerThrLysGlyLeuGlyLysThrValProLeuGlu 103
256 CCAAGGCAAGCAGACAGTGTCAAGAGTTTGGCAAAACTACCCCTGATCCT 305
104 SerGluPheValLysTrpArgAspValValValProCysGlyLysPr 120
306 TCAGCTAACATTAGT...CTGGATGGTGTAGACGTTCTCTTGGGACCGG 352
120 ovalProSerSerIleArgSerGluLeuMetTyrAsnGluTyrIleV 137
353 GATTTCATCTGGTGTGAATGACACCTCTTACATATATACGAGTATATG 402
137 alTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLysValArgPhe 153
403 TCTATGATATTGCTCAGGTAATCTGAAGTATCTGCTGAAACTGAAATTC 452
154 HisHisLys 156
453 AATTTTAAG 461

seq_name: gb_est2:BG042229

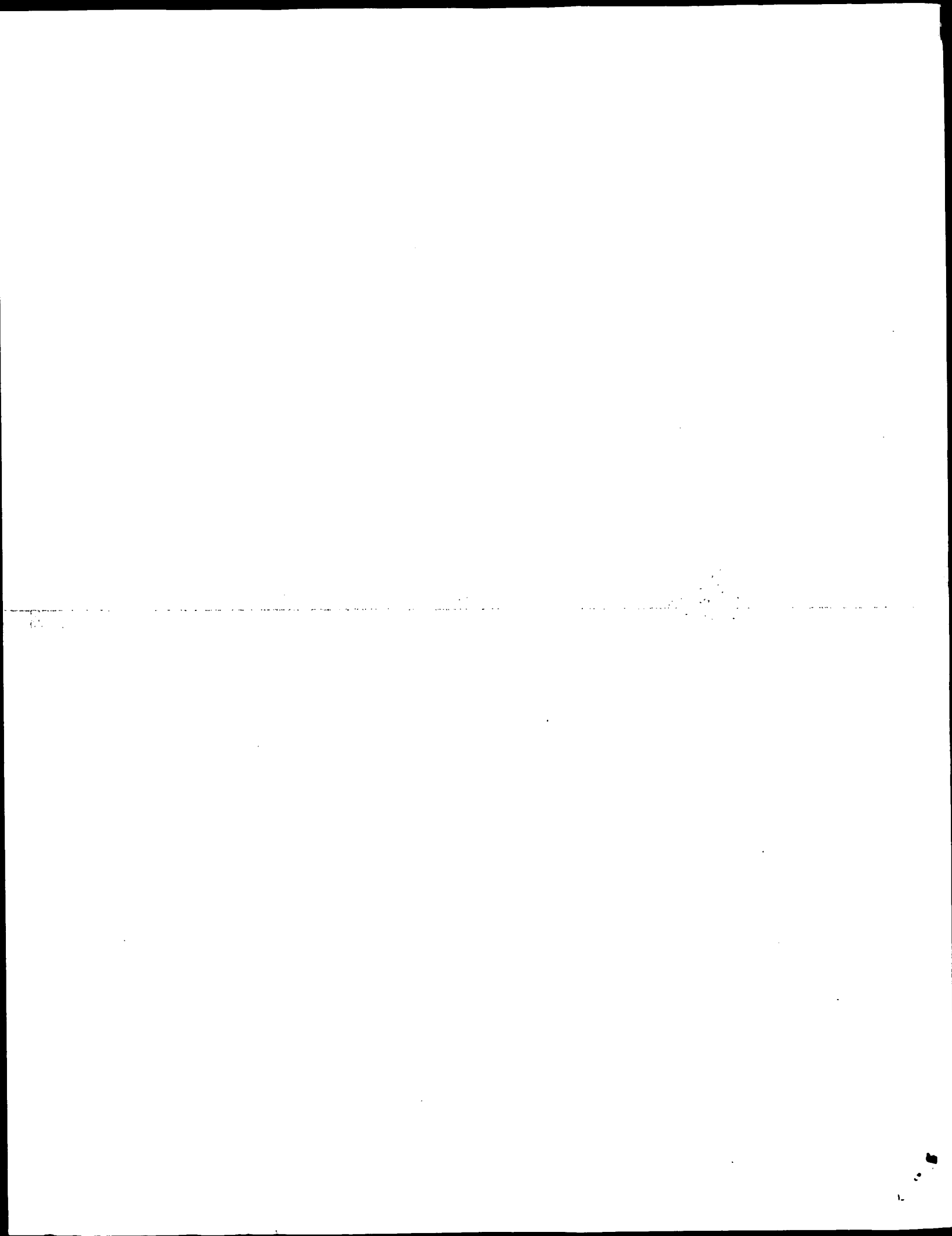
seq_documentation_block: 502 bp mRNA EST 31-JUL-2001
LOCUS BG042229 su93c12.y1 Gm-cl055 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl055-2064 5' similar to TR:Q9ZP54 Q9ZP54 POLY (ADP-RIBOSE)
POLYMERASE ;, mRNA sequence.
ACCESSION BG042229
VERSION BG042229.1 GI:12488706
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 502)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvill, V., Khanna
A., Boll, B., Marra, M., Hillier, L., Kucaba, F., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert Length: 429 Std Error: 0.00
High quality sequence stop: 419.
Location/Qualifiers
1..502
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl055-2064"
/clone_lib="Gm-cl055"
/tissue_type="Mature seed pods, greenhouse grown"

FEATURES
source

alignment_scores:
Quality: 503.50 Length: 153
Ratio: 3.757 Gaps: 1
Percent Similarity: 87.582 Percent Identity: 62.092
alignment_block:
US-09-236-995D-4 x BG675913
Align seg 1/1 to: BG675913 from: 1 to: 928

FEATURES
source

4 LeuLeuTrpHisGlySerArgLeuThrAsnPhleValGlyLeuSerG1 20
6 TTGCTGTGGCAGCGGTCCAGGACCACTTTGCTGGGATCTCTGTCCTCA 55
20 nGlyLeuArgIleAlaProGluAlaProValThrGlyTyrMetPheG 37
56 GGATCTTCGGATAGCCCGCTGAGCGCCGCTGAGCGGATCTCTGTCCTCA 105
37 lYlsGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCys 53
106 GTRAAAGGATCTATTTCGCTGACATGCTCCCAAGAGTGCCAACTACTGC 155
54 TyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSerGluValAl 70



OM of: US-09-236-995D-2 to: N_Geneseq_1101.* out_format : pfs
 Date: Mar 9, 2002 8:26 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
 -O=/cgn2.1/USPTO.spool/US09236995/runat_08032002.113149_25135/app_query.fasta_1.1273
 -DB=N_Geneseq_1101 -OFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPECI=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -GAPOP=6.000 -FGAPEXT=7.000 -START=1 -MATRIX=bysum62
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bysum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09236995 -CGNL_1_309 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-236-995D-2
 Query length: 982
 Database: N_Geneseq_1101.*
 Database sequences: 930621
 Database length: 428662619
 Search time (sec): 250.160000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|---|-----------|--------|---------|---------|---------------|
| /SIDS2/gcgdata/geneseq/NA1999.DAT:AAH89542 | | | 5036.00 | 8378.50 | 2949 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ60618 | | | 4740.00 | 7882.92 | 3212 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ60615 | | | 4666.50 | 7760.11 | 3211 |
| /SIDS2/gcgdata/geneseq/NA1996.DAT:AAZ13732 | | | 1594.50 | 2625.10 | 3793 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63954 | | | 1584.00 | 2610.08 | 3045 |
| /SIDS2/gcgdata/geneseq/NA1999.DAT:AAZ23799 | | | 1583.00 | 2606.55 | 3580 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63341 | | | 1518.50 | 2500.06 | 3200 |
| /SIDS2/gcgdata/geneseq/NA1995.DAT:AAZ04221 | | | 1237.00 | 2023.80 | 5345 |
| /SIDS2/gcgdata/geneseq/NA1995.DAT:AAZ04219 | | | 948.50 | 1549.66 | 2682 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63990 | | | 898.00 | 1462.87 | 3308 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63996 | | | 851.50 | 1393.77 | 3066 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44303 | | | 851.50 | 1392.08 | 1814 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ44287 | | | 851.50 | 1391.90 | 1843 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ632090 | | | 851.50 | 1391.90 | 1843 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ60617 | | | 851.50 | 1391.47 | 2147 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ60616 | | | 851.50 | 1390.14 | 2295 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAH14941 | | | 838.50 | 1367.65 | 1980 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63320 | | | 838.00 | 1368.52 | 864 |
| /SIDS2/gcgdata/geneseq/NA1995.DAT:AAZ04220 | | | 650.00 | 1074.78 | 1592 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44290 | | | 491.00 | 790.19 | 1740 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44291 | | | 490.50 | 790.41 | 1587 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44288 | | | 475.00 | 760.42 | 2265 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ44289 | | | 475.00 | 760.42 | 2265 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ80132 | | | 398.00 | 647.13 | 595 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85334 | | | 253.00 | 397.34 | 1143 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ45690 | | | 246.50 | 368.44 | 5490 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ92076 | | | 246.50 | 368.44 | 5490 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85318 | | | 225.00 | 358.96 | 550 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85319 | | | 225.00 | 358.96 | 550 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ66086 | | | 211.50 | 340.21 | 395 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAH42823 | | | 203.50 | 326.82 | 396 |
| /SIDS2/gcgdata/geneseq/NA1999.DAT:AAV82491 | | | 177.50 | 258.59 | 3417 |
| /SIDS2/gcgdata/geneseq/NA1996.DAT:AAZ9622 | | | 173.50 | 251.91 | 3417 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85306 | | | 161.50 | 256.07 | 416 |
| /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63838 | | | 157.50 | 223.41 | 3984 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ66826 | | | 157.50 | 222.98 | 4134 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ29627 | | | 157.50 | 222.98 | 4134 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ29632 | | | 157.50 | 222.03 | 4491 |
| /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ29633 | | | 157.50 | 221.61 | 4657 |

/SIDS2/gcgdata/geneseq/NA2001.DAT:AAH42831 + 151.00 238.55 3.3e-05 415
 /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ66824 + 135.00 184.95 0.0317 4296
 /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ85304 - 133.50 207.83 0.0017 472
 /SIDS2/gcgdata/geneseq/NA2000.DAT:AAZ77578 + 133.50 195.80 0.0079 1344
 /SIDS2/gcgdata/geneseq/NA2001.DAT:AAZ63919 + 133.50 186.68 0.0254 2971

seq_name: /SIDS2/gcgdata/geneseq/NA1999.DAT:AAH89542

seq_documentation_block:

ID: AAH89542 standard; cDNA; 2949 BP.

AC: AAH89542;

DT: 06-OCT-1999 (first entry)

DE: Maize poly ADP-ribose polymerase gene.

KW: PARP; poly ADP-ribose polymerase; metabolism; maize; transform; ss.

OS: Zea mays.

PH: Key Location/Qualifiers
 FT: CDS 1..2949

FT: /*tag= a

FT: /product= "PARP"

FT: /note= "Poly ADP-ribose polymerase"

FT: /transl_except= (pos:1453..1455, aa:Xaa)

FT: /transl_except= (pos:1495..1497, aa:Xaa)

FT: /transl_except= (pos:1504..1506, aa:Xaa)

FT: /transl_except= (pos:1534..1536, aa:Xaa)

FT: /transl_except= (pos:1537..1539, aa:Xaa)

FT: /transl_except= (pos:1561..1563, aa:Xaa)

FT: /transl_except= (pos:1582..1584, aa:Xaa)

FT: /transl_except= (pos:1588..1590, aa:Xaa)

FT: /transl_except= (pos:1594..1596, aa:Xaa)

FT: /transl_except= (pos:1975..1977, aa:Xaa)

FT: /transl_except= (pos:2011..2013, aa:Xaa)

FT: /transl_except= (pos:2014..2016, aa:Xaa)

FT: /transl_except= (pos:2026..2028, aa:Xaa)

FT: /transl_except= (pos:2047..2049, aa:Xaa)

FT: /transl_except= (pos:2050..2052, aa:Xaa)

FT: /transl_except= (pos:2053..2055, aa:Xaa)

FT: /transl_except= (pos:2056..2058, aa:Xaa)

FT: /transl_except= (pos:2077..2079, aa:Xaa)

FT: /transl_except= (pos:2095..2097, aa:Xaa)

FT: /transl_except= (pos:2107..2109, aa:Xaa)

FT: /transl_except= (pos:2116..2118, aa:Xaa)

FT: /transl_except= (pos:2119..2121, aa:Xaa)

FT: /transl_except= (pos:2122..2124, aa:Xaa)

FT: /transl_except= (pos:2125..2127, aa:Xaa)

FT: /transl_except= (pos:2128..2130, aa:Xaa)

FT: /note= "Xaa = unknown"

WO9937789-A1.

29-JUL-1999.

26-JAN-1999; 99WO-US01591.

27-JAN-1998; 98US-0072785.

(PION-) PIONEER HI-BRED INT INC.

Mahajan P, Zuo Z;

WPI; 1999-444613/37.

P-PSDB; AA228464.

New maize poly ADP-ribose polymerase gene useful in transforming plants to alter their metabolic state

Claim 1; Page 30-35; 48pp; English.

CC The sequence is the maize poly ADP-ribose polymerase (PARP) gene. The
CC sequence codes a PARP with 982 amino acids (AA28464). PARP is required
CC in the cell in most cases of DNA repair, recombination, rearrangement
CC and transposition. PARP gene and antisense gene can be used to transform
CC plant cells and alter the metabolic state of the transformed cell. This
CC is useful in enhancing disease resistance in plants and methods of
CC genetic transformation of plants. Plants transformed with either a sense
CC or antisense PARP nucleotide sequence may be utilized to increase
CC transformation frequency in plant cells. The enzyme also plays a role in
CC cellular stress, so may be beneficial for prevention of plant disease or
CC pathogen attack.
xx
50 Sequence 2949 BP; 895 A; 539 C; 732 G; 756 T; 27 other;

alignment_scores:
Quality: 5036.00 Length: 982
Ratio: 5.251 Gaps: 0
Percent Similarity: 97.658 Percent Identity: 97.658

alignment_block:
US-09-236-995d-2 x AAX89542 ..

Align seg 1/1 to: AAX89542 from: 1 to: 2949

1 MetAlaAlaProProLysAlaLirPlysAlaGluTyrAlaLysSerClyAr 17
1 ATGGCGCGCCGCCAAGGCGTGAAGCGGAGTATGCCAAGCTCGGGCG 50
17 gAlaSerCysLysSerCysArgSerProIleAlaLysAspGlnLeuArgL 34
51 GGCTCTGCTCAAGTCAATGCGGCTCCCTATCCCAAGGACAGCTCCGTC 100
34 euGlyLysMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrp 50
101 TTGGCAAGATGGTTACAGGCGTCACAGTTCCAGCGGCTTCATGCCGATGTG 150
51 AsnHisAlaArgCysIlePheSerLysLysAsnGlnIleLysSerValAs 67
151 AACCATGCCAGGTGCATCTTCACAGAGAACCCAGATAAATCCGTTGA 200
67 pAspValGluGlyIleAspAlaLeuArgTrpAspGlnGluLysIleA 84
201 CGATGTTGAAGGATAGATGACCTTATAGATGGGATGATCAAGAGAAGATAC 250
84 rGAsnTyrValGlySerAlaSerAlaGlyThrSerThrAlaAlaPro 100
251 GAAACTACGTTGGGAGTGCCTCAGCTGGTACAAGTTCTACAGCTGCTCCT 300
101 ProGluLysCysThrIleGluIleAlaProSerAlaArgThrSerCysAr 117
301 CCTGAGAAATGTACAATTGAGATTGCTCCATCTGCCCGTACTTCATGTAG 350
117 gArgCysSerGluLysIleThrLysGlySerValArgLeuSerAlaLysL 134
351 ACGTACGAGTCAAAAGATTACAAAGAGATCGGTCCTCTTCAGCTAAGC 400
134 euGluSerGluGlyProLysGlyIleProTyrHisAlaAsnCysPhe 150
401 TTGAGACTGAAGTCCCAAGGGTATACCATGGTATCATGCCAACTGTTTC 450
151 PheGluValSerProSerAlaThrValGluLysPheSerClyTrpAspTh 167
451 TTTGAGGTATCCCGCTGCAACTGTGTGAGAAGTTCTCAGGCTGGGATAC 500
167 rLeuSerAspGluAspLysArgThrMetLeuAspLeuValLysLysAspV 184
501 TTTGTCGATGAGGTCAAGAGACCATGCTCGATCTTTGTTAAAGAGATG 550
184 alGlyAsnAsnGlnAlaAsnLysGlySerLysArgLysLysSerGluAsn 200
551 TTGGCAACATGAACAAATAAGGGTTCCCAAGCGCAAGAAAGTGAAT 600

201 AspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrSerGluG 217
601 GATATTGATAGCTACAAATCCGCCAGGTTAGATGAAAGTACATCTGAAG 650
217 yThrValArgAsnLysGlyGlnLeuValAspProArgGlySerAsnThrS 234
651 TACAGTGCGAACAAAGGCGCAACTTGTAGACCCACGCTGGTTCCAACTA 700
234 erSerAlaAspIleGlnLeuLysLeuLysGlnSerAspThrLeuTtp 250
701 GTTCAGCTGATATCCCACTAAAGCTTAAGGAGCAAAAGTGACACACTTGG 750
251 LysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGluLeuArgAs 267
751 AAGTTAAAGGATGGACTTAAGACTCATGTATCGGCTGCTGAATTAAGGA 800
267 pMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArgHisLeuL 284
801 TATGCTTTGAGGCTAATGGGAGGATACATCAGGACCCAGAAAGGCACCTAT 850
284 euAspArgCysAlaAspGlyMetLeuPheGlyAlaLeuGlyProCysPro 300
851 TGGATCGCTGTGCGGATGGAATGCTATTGGAGCGCTGGCTTGCCTTGC 900
301 ValCysAlaAsnGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerG 317
901 GTCTGTGCTAATGGCATGTACTATTATAATGGTCAGTACCAATGCAGTGG 950
317 yAsnValSerGluTrpSerLysCysThrTyrSerAlaThrGluProValA 334
951 TAATGTGTACAGAGTGGTCCAAAGTGTACATCTCGCCACAGAACCTGTCC 1000
334 rGValLysLysLysTrpGlnIleProHisGlyThrLysAsnAspTyrLeu 350
1001 CGGTTAAGCAAGTGGCAAAATCCACATGCAACAAAGAAATGATTACCTT 1050
351 MetLysTrpPheLysSerGlnLysValLysLysProGluArgValLeuPr 367
1051 ATGAAGTGGTTCAAAATCTCAAAAGGTTAAGAAACCCAGAGAGGTTCTTCC 1100
367 oProMetSerProGluLysSerGlySerLysAlaThrGlnArgThrSerL 384
1101 ACCAATGTCACTCAGAAATCTGGAAGTAAAGCAACTCAGAGAACATCAT 1150
384 euLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValClyGln 400
1151 TGCTGCTCTCTAAAGGGTGGATAAAATTAAGGTTTCTGTTGTAGGACAA 1200
401 SerLysGluAlaAlaAsnGluTrpIleGluLysLeuLysLeuAlaGlyAl 417
1201 TCAAAAGAGCAGCAAAATGAGTGGATGGAAGCTCAAACTTGTGTGTGC 1250
417 aAsnPheTyrAlaArgValValLysAspIleAspCysLeuIleAlaCysG 434
1251 CAACCTTCTATGCCAGGTTGTCAAAAGATATTGATTGTTTAAATTCATGTG 1300
434 lyGluLeuAspAsnGluAsnAlaGluValArgLysAlaArgArgLeuLys 450
1301 GTGAGCTCGACAATGAAATGCTGAAGTCAGGAAGCAAGGAGGCTGAAG 1350
451 IleProIleValArgGluGlyTyrIleGlyGluCysValLysArgThrLy 467
1351 ATACCAATTTGAAGGAGGTTTACATTGGAGAAATGTTTAAAGAAACAAA 1400
467 sCysCysHisLeuIleCysIleAsnTrpAsnAlaLeuGluSerSerLysG 484
1401 ATGCTGCCATTTTGAATTTGATAAACTGGAATGCCCTTAGAGTCCCTCAAA 1450
484 ly***ThrValThrValLysValLysGlyArgSerAlaCysSer***Ser 500
1451 GCMGTACTGCTACTGTTAAAGTTAAGGCCGGAAGTCTGTTTCATYAAGT 1500
501 Pro***ValCysLysAsnThrAlaHisIlePro*****TrpGluLysHi 517


```

1501 CCTCYGTTTCAAGAAATACCTGCTACATTCCTTRAGRATGGGAAAAGCA 1550
|||||
517 sileGlnCys***LeuLysHisValLeuThr***His***ValCysThrG 534
|||||
1551 TATACATGCAACCTTAAACATGTTCTGACCTGNACACAGGTGGYACAG 1600
|||||
534 lyTyTyValLeuGlnlelleGluGlnAspAspGlySerGluCysTyr 550
|||||
1601 GCTACTATGCTACCTCAGATCATTTGAACAGGATGATGGCTCTGAGTGTAC 1650
|||||
551 ValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLys 567
|||||
1651 GTATTTCTGAAGTGGGACCGGTTGGGAGTGCAGAAAATTGGAGGGCAAAA 1700
|||||
567 sLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysArgLeuP 584
|||||
1701 ACTGAGGAGATGCAAAAACCTGAGGCAATCAAGGAATTCAAAAGATTAT 1750
|||||
584 heLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAsn 600
|||||
1751 TTCTTTGAGAAGACTGGGAACTCATGGGAAGCTTGGGAATCTAAAACCAAT 1800
|||||
601 PheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrGlyVa 617
|||||
1801 TTTCGGAAGCAGCCTGGGAGATTTTACCCACTTGATGTTGATTATGGTGT 1850
|||||
617 lLysLysAlaProLysArgLysAspIleSerGluMetLysSerSerLeuA 634
|||||
1851 TAAGAAAGCACCACAAACGGAACATATCATGTAAGTGAAGATTCCTTG 1900
|||||
634 laProGlnLeuLeuGluLeuMetLysMetLeuPheAsnValGluThrTyr 650
|||||
1901 CTCTCTCAATTGCTAGAACTCATGAAGATGCTTTTCAATGGGAGACATAT 1950
|||||
651 ArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetProLeuGl 667
|||||
1951 AGAGCTCTATGATGGAATTTGAAATTAATATGTCAGAAATGCCTCTTGG 2000
|||||
667 yLysLeuSer****AsnIleGlu***GlyPheGluAlaLeuThr**** 684
|||||
2001 GAAGCTAAGCAGGAGRAATATTCAGRAGGATTTGAAGCATTAACTKRG 2050
|||||
684 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu***Glu 700
|||||
2051 TACMGRATTTTATTGAAGGACACCGCTNATCAAGCACCTGGCTGTTGAG 2100
|||||
701 LysAla***LeuLeu*****PheSerLeuLeuLeuLeu 717
|||||
2101 AAAGCTNAATTTGTYTGSAGCMATSSYTTTTCACCTCTATCCCTTCT 2150
|||||
717 uPheIleLeuIleLeuTyrGlyMetArgMetIleSerTyrSerLysAlaL 734
|||||
2151 ATTCACTCTCATATTATACGGGATGAGGATGATTTTCATATTCNAAGCGCA 2200
|||||
734 ysMetLeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleValGly 750
|||||
2201 AAATGCTTGAAGCTCTGCAGGATATTGAAATTCCTTCAAGATAGTTGCC 2250
|||||
751 PheAspSerAspSerAspGluSerLeuAspAspLysTyrMetLysLeuHi 767
|||||
2251 TTCGATAGGACAGTAGTGAATCTCTTGATGATATAATATATAAAGCTTCA 2300
|||||
767 sCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeuIleG 784
|||||
2301 CTGTGACATCACCCGCTGGCTCAGCATAGTGAAGATTACAAGTTAATTG 2350
|||||
784 luGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSerLeu 800
|||||
2351 AGCAGTATCTCTCAACACACATGCTCTCTACTCACAAAGGACTGGTGTG 2400
|||||
801 GluLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTy 817
|||||

```

```

2401 GAACCTGGAGGAAGTTTTTTTCACTTGATGATGGAGAACTTAATAAGTA 2450
|||||
817 rSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlys 834
|||||
2451 CTCACGATATAAAATAATCTGCATAACAAGATGCTATTATGGCAGGTT 2500
|||||
834 exArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAla 850
|||||
2501 CAAGGTTCAGCAATTTTGTGGGAATCTTTAGTCAAGGGCTAAGAATTGCA 2550
|||||
851 ProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPh 867
|||||
2551 CTCTCTGAGGCACCTGTTACTGGCTATATGTTCCGCAAGGCTCTACTT 2600
|||||
867 eAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArgAsnA 884
|||||
2601 TCCAGATCTAGTAACCAAGAGCGCACATACATGTTATGTGGATAGAAVA 2650
|||||
884 snProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyr 900
|||||
2651 ATCTCTGATGTTTGTATGCTTCTTCTGAGGTGCTTTAGGACACATGTAT 2700
|||||
901 GluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysHisSe 917
|||||
2701 GAACTAAGAAAGCCACGTCCTCATGGACAAACCTCCAAGAGGGAAGCATTC 2750
|||||
917 rThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheValLysT 934
|||||
2751 GACCAAGGATTTAGCAAAACCGTCCACTGGAGTCAGAGTTTGTCAAGT 2800
|||||
934 rpArgAspAspValValValProCysGlyLysProValProSerSerIle 950
|||||
2801 GCAGGGATGATGCTGCTAGTTCCTCTCGGCAAGCGGTGCCATCATCAATT 2850
|||||
951 ArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSerGl 967
|||||
2851 AGGAGCTCTGAACCTCATGTACAATGAGTACATCGTCTACAACACATCCCA 2900
|||||
967 nValLysMetGlnPheLeuLeuLysValArgPheHisLysLysArg 982
|||||
2901 GGTGAAGATGCAGTTCTTCTGCAAGGTGCGTTTCCATCACAAGAGG 2946
|||||
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ60618
seq_documentation_block:
ID AAZ60618 standard; DNA; 3212 BP.
XX
AC AAZ60618;
XX
DT 16-MAY-2000 (first entry)
XX
DE DNA encoding the poly(ADP-ribose) polymerase ZAP2 protein of Zea mays;
XX
KW ZAP2; poly(ADP-ribose) polymerase; PARP; poly(ADP-ribose) transferase;
KW programmed cell death; apoptosis; growth rate; stress; cold; pathogen;
KW pest; drought; heat; fungi; nematode; seed-shatter; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 81..3023
FT /*tag= a
FT /product= "ZAP2 protein"
XX
PN WO200004173-A1.
XX
PD 27-JAN-2000.
XX
PF 12-JUL-1999; 99WO-EP04940.
XX
PR 17-JUL-1998; 98US-0118276.
XX
PA (PLB2 ) PLANT GENETIC SYSTEMS NV.

```

XX Babiychuk E, Kushnir S, De Block M;
 XX PI
 XX WPI: 2000-182436/16.
 XX P-PSDB: AAY68839.
 XX
 XX Modulating cell death, growth and stress resistance in eukaryotes,
 XX specifically plants, used, e.g. to impart fungus or nematode resistance
 XX
 XX Disclosure: Page 103-108; 126pp; English.
 XX
 XX The present sequence encodes the ZAP2 protein of Zea mays. This protein
 XX is a poly(ADP-ribose) polymerase (PARP) protein (also known as
 XX poly(ADP-ribose) transferase). The PARP enzyme is involved in programmed
 XX cell death or apoptosis, and is a nuclear enzyme. The ZAP2 polynucleotide
 XX sequences can be used for modulation of programmed cell death in
 XX eukaryotic cells. The method is used, specifically in plants, to induce,
 XX or protect against, programmed cell death, depending on the extent to
 XX which PARP activity is reduced. Reducing expression of endogenous ZAP
 XX class PARP only is also used to modulate programmed cell death, to
 XX increase growth rate and to produce plant cells that are more tolerant
 XX of stress (cold, chemical treatments, pathogens, pests, drought, heat,
 XX etc., or during transformation). Particular applications are generation
 XX of plants that are resistant to fungi or nematodes; are male or female
 XX sterile; or have better seed-shatter properties. The methods are also
 XX used to improve growth of transformed plant cells (and derived calli or
 XX complete plants).
 XX
 XX Sequence 3212 BP: 974 A; 600 C; 815 G; 823 T; 0 other;

alignment_scores:

Quality: 4740.00 Length: 985
 Ratio: 5.000 Gaps: 5
 Percent Similarity: 96.244 Percent Identity: 95.533

alignment_block:

US-09-236-995D-2 x AAZ60618

Align seg 1/1 to: AAZ60618 from: 1 to: 3212

1 MetAlaAlaProProLysAlaTrpLysAlaGluTyrAlaLysSerGlyAr 17
 81 ATGGCGCGCGCCAAAGCGTGGAAAGCGGAGTATGCCAAGCTCGGGCG 130
 17 gAlaSerCysLysSerCysArgSerProLleAlaLysAspGlnLeuArgL 34
 131 GGCCTCGTCAAGTCAATGCGGTCCTCCCTATCGCAAGGACCGAGCTCCGTC 180
 34 euGlyLysMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrp 50
 181 TTGGCAAGATGGTTACAGCGGTACAGTTCAGCGGCTTCATGCCGATGTGG 230
 51 AsnHisAlaArgCysLysPheSerTyrLysAsnGlnLleLysSerValas 67
 231 AACCATGCCAGGTGCATCTTCAGCAAGAAGAACCCAGATAAATCCGTGGA 280
 67 pAspValGluGlyLleAspAlaLeuArgTrpAspAspGlnGluLysLleA 84
 281 CGATGTTGAAGGATAGATGCACCTTAGATGGGATGATCAAGAGAAGATAC 330
 84 rGAsnTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaPro 100
 331 GAACTACGTTGGAGTGCTCAGCTGGTACAAAGTTCTACAGCTGCTCCT 380
 101 ProGluLysCysThrLleGluLleAlaProSerAlaArgThrSerCysAr 117
 381 CCTGAGAAATGTACAATTCAGATTGCTCCATCTGCCCGTACTTCATGTAG 430
 117 gArgCysSerGluLysLleThrLysGlySerValArgLeuSerAlaLysL 134
 431 ACGATGCGTGAAGGATACAAAGGATCGGTCGCTTTTCAGCTAAGC 480

134 euGluSerGluGlyProLysGlyLleProTrpTyrHisAlaAsnCysPhe 150
 481 TTGAGAGTGAAGGTCCCAAGGGTATACCATGGTATCATGCCAACTGTTTC 530
 151 PheGluValSerProSerAlaThrValGluLysPheSerGlyTrpAspTh 167
 531 TTTGAGGTATCCCGCTCTGCAACTGTGAGAAGTCTCTCAGGCTGGGATAC 580
 167 rLeuSerAspGluAspLysArgThrMetLeuAspLeuValLysLysAspV 184
 581 TTTGTCCGATCAGGATGAAGAACCATGCTCGATCTTGTAAAAAAGATG 630
 184 alGlyAsnAsnGluGlnAsnLysGlySerLysArgLysLysSerGluAsn 200
 631 TTGGCAACATGAACAAAAAAGGTTCGAAGGCCAAGAAAGTGAAT 680
 201 AspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrSerGluG 217
 681 GATATTGATAGCTACAAATCCGCCAGGTTAGATGAAAGTACATCTGAAGG 730
 217 yThrValArgAsnLysGlyGlnLeuValAspProArgGlySerAsnThrS 234
 731 TACAGTCGGAACAAAGGCAACTTTAGACCCACGCGGTTCACACTACTA 780
 234 erSerAlaAspIleGlnLysLysGluGlnSerAspThrLeuTrp 250
 781 GTTCAGCTGATATCCAACTTAAGCTTAAGGAGCAAGTACACACACTTTGG 830
 251 LysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGluLeuArgAs 267
 831 AAGTTAAAGGATGGACTTAAGACTCATGTATCGGCTGCTGAATTAAGGGA 880
 267 pMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArgHisLeuL 284
 881 TATGCTTGAGGCTAATGGGAGGATACATCAGGACCAGAAAGGACCATAT 930
 284 euAspArgCysAlaAspGlyMetLeuPheGlyAlaLeuGlyProCysPro 300
 931 TGGATCGCTGTGGGATGGAATGATATTGGAGCGCTGGGTCTGCCCCA 980
 301 ValCysAlaAsnGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerG 317
 981 GTCTGTCTAATGCGATGTACTATATATAGTGGTCAGTACCAATGCACTGG 1030
 317 yAsnValSerGluTrpSerLysCysThrTyrSerAlaThrGluProValA 334
 1031 TAATGTGTACAGTGGTCCAAAGTGTACATCTCTGCCACAGAACCTGTCC 1080
 334 rgValLysLysTrpGlnLleProHisGlyThrLysAsnAspTyrLeu 350
 1081 GCGTTAAGAGAAGTGGCAAAATCCACATGGACAAAGAAATGATTACCTT 1130
 351 MetLysTrpPheLysSerGlnLysValLysLysProGluArgValLeuPr 367
 1131 ATGAAGTGGTTCAAAATCTCAAAAGGTTAAGAACCCAGAGAGGGTCTTCC 1180
 367 oProMetSerProGluLysSerGlySerLysAlaThrGlnArgThrSerL 384
 1181 ACCAATGTCACTCAGAAATCTCGAAGTAAAGCAACTCAGAGAACATCAT 1230
 384 euLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValGlyCln 400
 1231 TGCTGTCTCTTAAAGGGTTGGATAAATTAAGGTTTCTGTGTAGGACAA 1280
 401 SerLysGluAlaAlaAsnGluTrpIleGluLysLeuLysLysLysL 417
 1281 TCAAAAGAGCAGCAAAATGAGTGGATTCAGAAAGCTCAAACTTGTGTGTG 1330
 417 aAsnPheTyrAlaArgValValLysAspIleAspCysLeuIleAlaCysG 434
 1331 CAACTTCTATGCCAGGTTGTCAAAGATATGATTGTTTAAATTCATGTG 1380

434 lYGlulLeuAspAsnGluAlaGluValArgLysAlaArgArgLeuLys 450
 1381 GTGAGTCGCAATGAAATGCTGAAGTCAGGAAGCAAGCAGGCTCAAG 1430
 451 ileProileValArgGluGlyTyrIleGlyGluCysVal.LysArgThrL 467
 1431 ATACCAATTTAGGGAGGGTTCATTTGGAGAATGTCTTAAAGAAACAA 1480
 467 yScysCysHisLeuIleCysIleAsn..TrpAsnAlaLeuGluSerSerL 483
 1481 AATGCTGCCATTTGATTTGATTAACACTAGAGAATGCTTAGAGTCTCTCAA 1530
 483 ySGly***ThrValThrValLysValLysGlyArgSerAlaCysSer*** 499
 1531 AAGCAGTACTGTCACCTGTTAAAGTTAAAGGGCGGAAGTGTCTTCATGAG 1580
 500 SerPro***ValCysLysAsnThrAlaHisIlePro*****TrpGluL 516
 1581 TCCTCTGGTTG...CAAGATACCTGCTACATCTCTTGA..AGATGGGAAA 1626
 516 sHisIleGlnCys***LeuLysHisValLeuThr***His***ValCys. 532
 1627 GCATATACAAATGCAACCTTAAACATGCTCTGACCTGCG...ACTAGTGTG 1673
 533 ThrGlyTyrTyrValLeuGlnIleIleGluGlnAspGlySerGluCy 549
 1674 AACAGCTACTATCTACTCCAGATCAATTAACAGGATGATGGGTCTCAGTG 1723
 549 sTyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyG 566
 1724 CTACGTATTTCTGTAAGTGGGACGGTGGGAGTGAGAAATTTGAGGGC 1773
 566 lnLysLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysArg 582
 1774 AAAAAGCTGGAGGAGATGTCAAAGCTAGAGCAATCAAGGAATTCAAAAGA 1823
 583 LeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysTh 599
 1824 TTATTTCTTGAAGAAGCTGGAAGCTCATGGGAAGCTTGGGAATGTAAAG 1873
 599 rAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrG 616
 1874 CAATTTTCGAGACAGCTGGAGATTTTACCACCTTGATGTGTGATATG 1923
 616 lyValLysLysAlaProLysArgLysAspIleSerGluMetLysSerSer 632
 1924 GTGTTAAGAAAGCAACCAACGGAAGATATCAGTGAATGAAAGTTCT 1973
 633 LeuAlaProGlnLeuGluLeuMetLysMetLeuPheAsnValGluTh 649
 1974 CTGTCTCTCAATGCTTGAAGCTCATGAAGATGCTTTTCAATGTGGAGAC 2023
 649 rTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetProL 666
 2024 ATATAGAGCTGCTATGATGAATTTGAAATTAATATGTCAGAAATGCCTC 2073
 666 euGlyLysLeuSer*****AsnIleGlu***GlyPheGluAlaLeuThr 682
 2074 TTGGGAAGCTAAGCAAGCAAAATATTTGAGAAAGGATTTGAAAGCATTAAC 2123
 683 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu** 699
 2124 GAGATACAGATTTTATT..GAAGACACCGCTGATCAAGCATGGCTG..TT 2171
 699 *GluLysAla***LeuLeu*****PheSerLeuLeuSerL 716
 2172 AGAGAAAGCTTAATTTGCTGCGAGCAATCGCTTTTCACTCTTATCCC 2221
 716 euLeuPheIleLeuTyrGlyMetArgMetIleSerTyrSerLys 732
 2222 TCTATTCTATCTCATATATACGGGATGAGCATGATTG..ATGATCAA 2270
 733 AlaLysMetLeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleVa 749

2271 GCGAAATCTTTGAAGCTCTCGAGGATATTGAATTTGCTTCAAGATAGT 2320
 749 lGlyPheAspSerAspSerAspGluSerLeuAspAspLysTyrMetLysL 766
 2321 TGGCTTCGATAGCAGCAGTGAATCTCTTGTATGATAAATATATGAAC 2370
 766 euHisCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeu 782
 2371 TTCACCTGTGACATCACCCCGCTGGCTCAGATAGTAGAAGATTACAAAGTTA 2420
 783 lLeGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSe 799
 2421 ATTGAGCAGTATCTCTCAACACACATGCTCTACTCACAGGACTGGTC 2470
 799 rLeuGluLeuGluValPheSerLeuAspArgAspGlyGluLeuAsnL 816
 2471 GCTGAACTGGAGGAAGTTTTTTCACCTTTCATCGAGATGAGAACTTAATA 2520
 816 ySTyrSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHis 832
 2521 AGTACTCAAGATATAAAATTAATCTGCATACAAAGATGCTATTATGGCAC 2570
 833 GlySerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIl 849
 2571 GGTTCAGGTTGACGAATTTTGTGGGAATTTCTAGTCAAGGCTTAAGAAT 2620
 849 eAlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuT 866
 2621 TGCACCTCTGAGGCACCTGTTACTGGTATATGTTGGCAAGGCTCTCT 2670
 866 yrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArg 882
 2671 ACTTGCAGATCTAGTAGCAAGAGCGGCAATACTGTTATGTGGTAGG 2720
 883 AsnAsnProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMe 899
 2721 AATAATCTGTAGGTTGATGCTCTTCTGAGGTGCTTTTAGGAGACAT 2770
 899 tTyrGluLeuLysLysAlaThrSerMetAspLysProArgGlyLysH 916
 2771 GTATGAACATAAGAAAGCCACGCTCCATGGACAAACCTCCAAGAGGGAAG 2820
 916 isSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 932
 2821 ATTCGACCAAGGATTAAGCAAAACCTGCCACTGGAGTCAGAGTTTGTG 2870
 933 LysTrpArgAspValValProCysGlyLysProValProSerSe 949
 2871 AAGTGGAGGATGATGCTGTAGTTCCCTCGGCAAGCCGCTGCATCATC 2920
 949 rIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrS 966
 2921 AATTAGAGCTCTCGAATCATGATGATGATGATGATGATGATGATGATGAT 2970
 966 erGlnValLysMetGlnPheLeuLysValArgPheHisHisLysArg 982
 2971 CCCAGGTGAAGTGCAGTCTTCTGCTGAAGGTGCTTTCCATCACAAGAGG 3020

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ60615

seq_documentation_block:

ID AAZ60615 standard; DNA; 3211 BP.

XX
AC
AAZ60615;

XX
DT 16-MAY-2000 (first entry)

XX
DE DNA encoding the poly(ADP-ribose) polymerase ZAP1 protein of Zea mays.
XX
KW ZAP1; poly(ADP-ribose) polymerase; PARP; poly(ADP-ribose) transferase;
KW programmed cell death; apoptosis; growth rate; stress; cold; pathogen;
KW pest; drought; heat; fungi; nematode; seed-shatter; ss.

XX OS Zea mays.
XX FH key Location/Qualifiers
XX FT CDS 113..3022
XX FT /*tag= a
XX FT /product= "ZAP1 protein"
XX PN WO200004173-A1.
XX PD 27-JAN-2000.
XX PF 12-JUL-1999; 99WO-EP04940.
XX PR 17-JUL-1998; 98US-0118276.
XX PA (PLBZ) PLANT GENETIC SYSTEMS NV.
XX PI Babyichuk E, Kushnir S, De Block M;
XX P-PSDB: AAY68833.
XX DR WPI; 2000-182436/16.
XX PT Modulating cell death, growth and stress resistance in eukaryotes,
XX PT specifically plants, used, e.g. to impart fungus or nematode resistance
XX PS Example 1; Page 79-84; 125pp; English.
XX CC The present sequence encodes the ZAP1 protein of Zea mays. This protein
CC is a poly(ADP-ribose) polymerase (PARP) protein (also known as
CC poly(ADP-ribose) transferase). The PARP enzyme is involved in programmed
CC cell death or apoptosis, and is a nuclear enzyme. The ZAP1 polynucleotide
CC sequences can be used for modulation of programmed cell death in
CC eukaryotic cells. The method is used, specifically in plants, to induce,
CC or protect against, programmed cell death, depending on the extent to
CC which PARP activity is reduced. Reducing expression of endogenous ZAP
CC class PARP only is also used to modulate programmed cell death, to
CC increase growth rate and to produce plant cells that are more tolerant
CC of stress (cold, chemical treatments, pathogens, pests, drought, heat,
CC etc., or during transformation). Particular applications are generation
CC of plants that are resistant to fungi or nematodes; are male or female
CC sterile; or have better seed-shatter properties. The methods are also
CC used to improve growth of transformed plant cells (and derived calli or
CC complete plants).
XX SQ Sequence 3211 BP; 968 A; 604 C; 813 G; 826 T; 0 other;

alignment_scores:
Quality: 4666.50 Length: 985
Ratio: 4.980 Gaps: 6
Percent Similarity: 95.127 Percent Identity: 94.416

alignment_block:
US-09-236-995D-2 x AAZ60615 ..

Align seg 1/1 to: AAZ60615 from: 1 to: 3211

1 MetAlaAlaProProLysAlaTrpLysAlaGluTyrAlaLysSerGlyAr 17
113 ATGGCGCGCGCGCAAGAGCGTGGAGCGGAGTATGCCAAGTCTGGGCG 162
17 qAlaSerCysLysSerCysArgSerProLysAlaLysAspGlnLeuArgL 34
163 GGCCCTCGTGCAGTATCGCCGGTCCCTATCGCCAGGACCGAGTCCGTC 212
34 euGlyLysMetValGlnAlaSerGlnPheAspGlyPheMetProMetTrp 50
213 TTGGCAAGATGGTTTCAGCGCTCACAGTTCGACGGCTTCATGCCGATCTGG 262
51 AsnHisAlaArgCysIlePheSerLysLysAsnGlnIleLysSerValAs 67
|||||

263 AACCATGCC.....AGCGTTGA 279
67 pAspValGluGlyIleAspAlaLeuArgTrpAspAspGlnGluLysIleA 84
|||||
280 CGATGTTGAAGGGATAGATGCACCTTAGATGGGATGATCAAGAGAAGATAC 329
84 rgAsnTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaPro 100
|||||
330 GAAACACTGTTGGAGTGCCTCAGCTGGTACAAGTTCTACAGCTGCTCCT 379
101 ProGluLysCysThrIleGluIleAlaProSerAlaArgThrSerCysAr 117
|||||
380 CCTCAGAAATGTACAATTTGAGATTGCTTCCATCTGCCCGTACTTTCATGTAG 429
117 gArGcysSerGluLysTleThrLysGlySerValArgLeuSerAlaLysL 134
|||||
430 ACGATGCGTGAAGAGATTAACAAAGAGATGCGTCCGCTCTTTCAGCTAAGC 479
134 euGluSerGluGlyProLysGlyIleProTrpTyrHisAlaAsnCysPhe 150
|||||
480 TTGAGAGTGAAGGTCCCAAGGTATACCATGGTATCATGCCAACTGTTC 529
151 PheGluValSerProSerAlaThrValGluLysPheSerGlyTrpAspTh 167
|||||
530 TTTGAGGTATCCCGCTCTGCAACTGTTGAGAAGTCTCAGCGCTGGGATAC 579
167 rLeuSerAspGluAspLysArgThrMetLeuAspLeuValLysLysAspV 184
|||||
580 TTTGTCGATGAGGATTAAGAGAACCATGCTCGATCTTGTAAAAAAGATG 629
184 alGlyAsnAsnGluGlnAsnLysGlySerLysArgLysLysSerGluAsp 200
|||||
630 TTGGCACAATGAACAAATAAGGGTTCCAAAGCGCAAGAAAGTGAATAAT 679
201 AspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrSerGluGl 217
|||||
680 GATATTGATGACTACAAATCGCCAGGTAGATGAAAGTACATCTGAAGG 729
217 YThrValArgAsnLysGlyGlnLeuValAspProArgGlySerAsnThrS 234
730 TACAGTGCAGAAACAAAGGGCAACTGTAGACCCACGTTGTTCCAACTACTA 779
234 erSerAlaAspIleGlnLeuLysLysGluGlnSerAspThrLeuTrp 250
|||||
780 GTTCAGCTGATATCCAACTAAAGCTTAAGGAGCAAGAGTGACACACTTTGG 829
251 LysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGluLeuArgAs 267
|||||
830 AAGTTAAAGGATGGACTTAAGACTCATGTATCGGCTGCTGAATTAAGGGA 879
267 pMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArgHisLeuL 284
|||||
880 TATGCTTGAGGCTAATGGGAGGATACATCAGGACCCAGAAAGGCACCTAT 929
284 euAspArgCysAlaAspGlyMetLeuPheGlyAlaLeuGlyProCysPro 300
|||||
930 TGGATCGCTGTGGGATGGAATGATATTTGGAGCGCTGGGTCCTTGCCCA 979
301 ValCysAlaAsnGlyMetTyrTyrTyrAsnGlyGlnTyrGlnCysSerGl 317
|||||
980 GTCTGTGCTAATGGCATCTACTATTATTAATGTCAGTACCAATGCAGTGG 1029
317 YasnValSerGluTrpSerLysCysThrTyrSerAlaThrGluProVala 334
|||||
1030 TAAATGTGTGAGAGTGGTCCAAAGTGTACATCTCTGCCACAGAACTGTGCC 1079
334 qValLysLysLysTrpGlnIleProHisGlyThrLysAsnAspTyrLeu 350
|||||
1080 GCGTTAAGAAAGAGTGGCAATTCACATGGAACAAAGAATGATTACCTT 1129
351 MetLysTrpPheLysSerGlnLysValLysLysProGluArgValLeuPr 367
|||||
1130 ATGAAGTGTTCAAATCTCAAAAGGTTAAGAAACCCAGAGAGGGTCTCTTCC 1179

367 oProMetSerProGluLysSerGlySerLysAlaThrGlnArgThrSerL 384
|||||
1180 ACCAATGTCACCTGAGAAATCTGCAAGTAAAGCAACTCAGAGAACATCAT 1229
384 euLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValGlyGln 400
|||||
1230 TGGTCTCTCTAAAGGGTTGGATAAATTAAGGTTTTCTGTGTAGACAA 1279
401 SerLysGluAlaAlaAsnGluTrpIleGluLysLeuLysLeuAlaGlyAl 417
|||||
1280 TCAAAAGACGACCAATGAGTGGATGGAAGCTCAAACTTGCTGGTGC 1329
417 aAsnPheTyrAlaArgValValLysAspIleAspCysLeuIleAlaCysG 434
|||||
1330 CAACCTCTATGCCAGGGTTGTCAAGATATATGTTTAAATGTCATGTG 1379
434 lyGluLeuAspAsnGluAsnAlaGluValArgLysAlaArgLeuLys 450
|||||
1380 GTGAGCTCGCAATGAATGCTGAAGTCAGGAAGCAAGGAGGCTGAAG 1429
451 IleProIleValArgGluGlyTyrIleGlyGluCysVal.LysArgThrL 467
|||||
1430 ATACCAATGCTAAGGAGGGTTACATTGAGAAATGTTTAAAGAACAA 1479
467 yScysCysHisLeuIleCysIleAsn..TrpAsnAlaLeuGluSerSerL 483
|||||
1480 AATGCTGCCATTTGATTTGTATAACTAGAGAAATGCTTAGAGTCTCTCA 1529
483 yGlyLysThrValThrValLysValLysGlyArgSerAlaCysSer*** 499
|||||
1530 AAGCGAGTACTGTCACCTGTTAAAGTTAAGGGCCGAGTCTGTCATGAG 1579
500 SerPro***ValCysLysAsnThrAlaHisIlePro****TrpGluLys 516
|||||
1580 TCCTCTGGTTG...CAAGATACTGTCACATTCITGA.AGATGGGAAA 1625
516 sHisIleGlnCys***LeuLysHisValLeuThr***His***ValCys. 532
|||||
1626 GCATATACAACTGACCTTAAACATGTCACCTGGC...ACTAGTGTG 1672
533 ThrGlyTyrTyrValLeuGlnIleIleGluGlnAspAspGlySerGluCy 549
|||||
1673 AACAGCTACTATGACTCCAGATCATTTGAACAGGATGATGGTCTGAGTG 1722
549 sTyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlyG 566
|||||
1723 CTACGTATTTCTGTAAGTGGGGACGGGTTGGAGTGAAGAAATGGAGGCG 1772
566 loLysLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysArg 582
|||||
1773 AAAAAGTGGAGAGATGTCAAAACCTGAGGCATCAAGGAATTCAAAAGA 1822
583 LeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysTh 599
|||||
1823 TTATTTCTTTGAGAGACTGGAAACTCATGGGAAGCTTGGGAATGTAAAC 1872
599 rAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrG 616
|||||
1873 CAATTTTCGGAGACGCTGGGAGATTTTACCCACTTGATGTTGATTATG 1922
616 lyValLysLysAlaProLysArgLysAspIleSerGluMetLysSerSer 632
|||||
1923 GTGTTTAAAGACCAACCAACGGAAGATATACAGTGAATGAAAGTTCT 1972
633 LeuAlaProGlnLeuLeuGluLeuMetLysMetLeuPheAsnValGluTh 649
|||||
1973 CTTGCTCTCAATTCCTAGAACTCATGAAGATGCTTTTCAATGTGGAGAC 2022
649 rTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetProL 666
|||||
2023 ATATAGAGCTGCTATGATGGAATTTGAAATTAATATGTCAGAAATGCTC 2072

666 euGlyLysLeuSer*****AsnIleGlu***GlyPheGluAlaLeuThr 682
|||||
2073 TTGGGAGCTAAGCAAGGAAATATTTGAGAAAGGATTTGAGCACTTAAC 2122
683 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu** 699
|||||
2123 GAGATACAGAAATTTATT.GAAGGACACCGCTGATCAAGCACTGGCTG.TT 2170
699 *GluLysAla***LeuLeu*****PheSerLeuLeuSerL 716
|||||
2171 AGAGAAAGCTTAATTTGCTGCGAGCAATCGCTTTTCACTCTTATCCC 2220
716 euLeuPheIleLeuIleLeuTyrGlyMetArgMetIleSerTyrSerLys 732
|||||
2221 TTCTATTCATCTCATATTTATACGGGATGAGGATGATTG.ATGATCAA 2269
733 AlaLysMetLeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleVa 749
|||||
2270 GCGAAATGCTTGAAGCTCTGCAGGATATTGAAATTTGCTTCAAGATAGT 2319
749 lGlyPheAspSerAspSerAspGluSerLeuAspLysTyrMetLysL 766
|||||
2320 TGGCTTCGATACGACGATGATGAATCTCTTGATGATAAATATATGAAC 2369
766 euHisCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeu 782
|||||
2370 TTCCTGTGACATCACCCCGCTGGCTCAGCATAGTGAAGATTACAAGTTA 2419
783 IleGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSe 799
|||||
2420 ATTGAGCAGTATCTCTCAACACACATGCTCTACTCAAGGACTGGTC 2469
799 rLeuGluLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnL 816
|||||
2470 CTTGGAACCTGGGAGAGTTTTCCTGATCGAGATGGGAACTTAATA 2519
816 yTyrSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHis 832
|||||
2520 AGTACTCAAGATATAAAAAATAATCTGCATAACAAGATGCTATTATGGCAC 2569
833 GlySerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgII 849
|||||
2570 GGTTCAAGTTGACGAATTTTGTGGGAATTTCTTAGTCAAGGGCTAAGAAT 2619
849 eAlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeu 866
|||||
2620 TGCACCTCTCTGAGGACCTGTTACTGCTATATGTTGGCAAGGCCCTCT 2669
866 yrPheAlaAspLeuValSerLysSerAlaGlnTyrCystyrValAspArg 882
|||||
2670 ACTTTCAGATCTAGTAAGCAAGAGCGCACAAATACTGTTATGTGGATAG 2719
883 AsnAsnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMe 899
|||||
2720 AATACTCTGTAGTTTGTATGCTTCTTCTGAGGTTGCTTTTAGGACAT 2769
899 tTyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysH 916
|||||
2770 GTATGAACATAAGAAAGCCAGCTCCATGGACAAACCTCCCAAGAGGAGAC 2819
916 isSerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 932
|||||
2820 ATTGCAACCAAGGATTAGGCAAAACCGCTGCACTGGAGTCAGAGTTTGTG 2869
933 LysTrpArgAspValValValProCysGlyLysProValProSerSe 949
|||||
2870 AGTGGAGGATGATGCTGTTAGTTCCCTGGGCAAGCCGCTGCATCATC 2919
949 rIleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrS 966
|||||
2920 AATTAGGAGCTCTGAATCATGTAACATGAGTACATGCTGTACACACAT 2969
966 erGlnValLysMetGlnPheLeuLysValArgPheHisLysArg 982

2666 ATTGCTGTGGCAGCGGTCCAGGACCAACCACTTTGCTGGGATCTCTGTGCC 2715
 845 InGlyLeuArgIleAlaProProGluAlaProValThrGlyTyrMetPhe 861
 2716 AGGGTCTTCGATAGCCCCCGCTGAAGCGCGGTGACAGGTACATGTTT 2765
 862 GlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCy 878
 2766 GCTAAGGAGTCTATTTCGCTGACATGCTCCCAAGAGTGCCAACTACTG 2815
 878 sTyrValAspArgAsnAsnProValGlyLeuMetLeuLeuSerGluVala 895
 2816 CCATACGCTCTCAGGAGACCCCAATAGGCTTAATCCTGTGGGAGAAGTTG 2865
 895 laLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMetAspLysPro 911
 2866 CCCTTGGAAACATGATGACTGAAGCAGCTTCACATATCAGCAAGTTA 2915
 912 ProArgGlyLysHisSerThrLysGlyLeuGlyLysThrValProLeuG1 928
 2916 CCCAAGGCGACGACAGTGTCAAAGTTTGGCAAACTACCCCTGTATCC 2965
 928 uSerGluPheValLysTyrArgAspAspValValValProCysGlyLysP 945
 2966 TTCACCTAACATTAAGT...CTGGATGGGTAGACGTTCTCTTGGGACCG 3012
 945 roValProSerSerIleArgSerSerGluLeuMetTyrAsnGluTyrIle 961
 3013 GGATTTTCATCTGGTGTGAATGACACCTCTCTACTATATAACGAGTACAT 3062
 962 ValTyrAsnThrSerGlnValLysMetGlnPheLeuLeuLysValArgPh 978
 3063 GTCATGATGATGTCACAGTAAATCTGAAGTATCTGCTGAAACTGAAATT 3112
 978 ehIshisLys 981
 3113 CAATTTTAAG 3122

seq_name: /SIDS2/qcgdata/geneseq/geneseqn/NA2001.DAT.AAF63954

seq_documentation_block:

ID AAF63954 standard; DNA; 3045 BP.

XX AAF63954;

XX 05-APR-2001 (first entry)

DE Human tankyrase2 related coding sequence SEQ ID NO: 136.

XX Human; tankyrase2; TANK2; TRF1; telomere; cancer; neoplasia; aging;
 KW inflammatory disorder; ds.

XX Homo sapiens.

XX WO200100849-A1.

XX 04-JAN-2001.

XX 28-JUN-2000; 2000WO-US17827.

XX 29-JUN-1999; 99US-0141582.

XX (ICOS-) ICOS CORP.

XX Christenson E, Demaggio AJ, Goldman PS, McElligott DL;

XX WPI; 2001-102896/11.

DR P-PSDB; AAB66296.

XX New tankyrase2 polypeptides, useful for treating conditions mediated by
 PT poly(adenosine diphosphate-ribose) polymerase activity e.g. cancers,
 PT inflammatory and autoimmune disorders -

XX

PS Example 2; Page 203-207; 242pp; English.

XX The present invention provides the protein and coding sequence for the
 CC human tankyrase2 protein. This is found in two different versions,
 CC designated TANK2-LONG and TANK2-SHORT. Tankyrase2 has
 CC polyADP-ribosylation activity and is involved in the modification of
 CC TRF1, which is a telomere-specific binding protein. The regulation of
 CC telomere length, in which TRF1 has a role, is linked to ageing and
 CC cancer. The sequences are useful in the treatment of cancers and
 CC inflammatory disorders.

XX Sequence 3045 BP; 863 A; 710 C; 851 G; 621 T; 0 other;

alignment_scores:

Quality: 1584.00 Length: 1060

Ratio: 2.289 Gaps: 36

Percent Similarity: 65.283 Percent Identity: 36.226

alignment_block:

US-09-236-995D-2 x AAF63954 ..

Align seg 1/1 to: AAF63954 from: 1 to: 3045

6 LysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLysSe 22
 19 AAGCTCTATCGAGTCGAGTACGCCAAGAGCGGCGCGCTCTTGCAGAA 68
 22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetValG 39
 69 ATCCAGCAGAGCATCCCAAGGACTCGCTCCGGATGGCCATCATGGTGC 118
 39 InAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
 119 AGTCGCCCATGTTGATGGAAGTCCACACTGGTACCACCTCTCCTCC 168
 56 IlePheSerLysLysAsnGlnLysSerValAsp...AspValGluG1 71
 169 TTCTGGAAGGTGGGCCACTCCATCCGCGCACCTGACGTTGAGGTGGATGG 218
 71 YlleAspAlaLeuArgTrpAspAspGlnGluLysIleArgAsnTyrVal. 87
 219 GTTCTCTGAGCTTCGGTGGATGACCAGCAGAAAGTCAAGAACAGACGCG 268
 88GlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
 269 AAGCTGGAGGAGTGACAGGCAAGGCCAGGATGGAATTGGTAGCAAGGCA 318
 102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
 319 GAGAAGACTCTGGGTGACTTTGCGCAGAGATGTCACAGTCCCAACAGAAG 368
 114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLeuS 131
 369 TACGTGCAAGGGGTGTATGGAGAAGATAGAAAGGCCGCGGCTGTCT 418
 131 erAlaLysLeu...GluSerGluGlyProLys...GlyIle.....Pro 143
 419 CCAAGAAGATGGTGGACCCGAGAGCCACAGCTAGGCATGATTGACCGC 468
 144 TrpTyrHisAlaAsnCysPhePheGlu.....ValSe 154
 469 TGTACCATCCAGGCTGCTTTGTCAAGAACAGGAGGAGTGGGTTTCGC 518
 154 rProSerAlaThrValGluLysPheSerGlyTrpAspThrLeuSerAspG 171
 519 GCCCGAGTACAGTGGCAGCTCAGGCTCAAGGGCTTCAGCCTCCTTGTACAG 568
 171 luAspLysArgThrMetLeuAspLeuValLysLysAspValGlyAsnAsn 187
 569 AGGATAA.....GAAGCCTGAAGAAGCAGCTCCAGGAGTC 606
 188 GluGlnAsnLysGlySerLysArgLysLysSerGluAsnAspIleAspSe 204

607 AAG.....AGTGAAGAAAGAGAAAGAGCGGATGAGGTGGATGG 644
204 rTyrLysSerAlaArgLeuAspGluSerThrSerGluGlyThrValArgA 221
645 A.....GTGGATCAAGTGGCGAAGAGAAATCTAAAGAG 679
221 snLysGlyGlnLeuValAspProArgGlySerAsnThrSerSerAlaAsp 237
680 AAAAGACAG.....GATAGTAAG 699
238 IleGlnLeuLysLeuLysGluGlnSerAspThrLeuTrpLysLeuLysAs 254
700 CTGAAAAAGCCCTAAAGGCTCAGAAGCAGCTGATCTGGAACATCAAGGA 749
254 pGlyLeuLysThrHisValSerAlaAlaGluLeuArgAspMetLeuGluA 271
750 CGAGCTAAAGAAAGTGTTCACCTAATGACCTGAAGGAGCTACTCATCT 799
271 laAsnGlyGlnAspThrSerGlyProGluArgHisLeuLeuAspArgCys 287
800 TCAACAGCAGCAAGTGCCTCTGGGAGTTCGGCGATCTTGGACCGAGTA 849
288 AlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAlaAs 304
850 GCTGATGGATGTTGTTGGTCCCTTCCTTCCTCGGAGGAATGCTCGGG 899
304 nGlyMetTyrTyrTyrAsnGlyGlnTyrGlnCysSerGlyAsnValSerG 321
900 TCAGCTGGTCTTCAAGAGCGATGCCTATTACTGCACTGGGAGCTCAGTG 949
321 luTrpSerLysCysThrTyrSerAlaThrGluProValArgValLysLys 337
950 CTTGGACCAAGTGTATGGTCAAGACAGACACACCAACCGG.....AAG 993
338 LysTrpGlnIleProHisGlyThrLysAsnAspTyrLeuMetLysTrpPh 354
994 GAGTGGTAAACCCA.....AAGNAATCCGAGAAATCTTACCT 1034
354 eLysSerGlnLysValLysLysProGluArgValLeuProMetSer. 370
1035 CAAGAAATGAAGGTATAAAGCAGGACCGTATATTCCCGCCAGAAACCA 1084
371ProGluLysSerGlySerLysAla 378
1085 GGCCTCCGTGGCGCCAGCGCTCCGCCCTCCACAGCCTCGGCTCCTGCT 1134
379 ThrGlnArgThrSerLeuSerLysGlyLeuAspLysLeuArgPh 395
1135 GCTGTGAACCTCTCTGCTTACAGCATAGCCATTATCCACATGAAGAT 1184
395 eSerValValGly.....GlnSerLysGluAlaAlaAsnGluTrpI 409
1185 CCGTACTCGGGAAGTGTCCCGGAACAGGATGAAGTGAAGGCCAIGA 1234
409 leGluLysLeu.....LysLeuAlaGlyAlaAsnPheTyrAlaArgVal 423
1235 TTGAAAATCGGGGGGAAGTTGAGGGGAGCGGCCAACCAAGGCTTCCCTG 1284
424 ValLysAspIleAspCysLeuIleAlaCysGlyGluLeuAspAsnGluAs 440
1285TGCAATCAGCCCAAAAGAGGTTGGAAAAGATGAA 1319
440 nAlaGluValArgLysAlaArgLeuLysIleProIleValArgGluG 457
1320 TAAGAAGATGGAGGAAGTAAAGGAAGCAACATCCGAGTTGTGTGAGG 1369
457 lyTyrlleGlyCysValLysArgThrLys..... 467
1370 ACTTCCTCCAGAGCTCTCCGCCCTCCACCAAGAGCCTTCAGGATTTGTC 1419
468 CysCysHisLeuIleCysIleAsnTrpAsnAlaLeuGluSerLysGI 484
..... 499

1420 TTAGCGCATCTTCTGCC...CCTTGGGGGCGAGAGTGAAGCGAGAGCC 1466
484 y***ThrValThrValLysValLysGlyArgSerAlaCysSer***SerP 501
1467 TGTTCGAAGTT...GTGGCCCAAGAGGAAGTCAAGGGCTGCGTCTCCA 1513
501 to***ValCysLysAsnThrAlaHisIlePro****Trp..... 514
1514 AA.....AAAAGCAAGGCCAGGTCAGAGGAAGGTATCAACAA 1554
515 ...GluLysHisIleGlnCys***LeuLys..... 523
1555 TCTGAAAAGAGATTAATACTTTAAAGGAGGAGCAGCTGGATCC 1604
524HisValLeu..... 526
1605 TGATTCTGGACTGGAACTCTGGCATCTCTGAGAAAGGTGGGAGG 1654
527Thr***His***Val.....CysThrGly 534
1655 TCTTCAGTGCACCTTGGCCTGGTGGACATCGTTAAAGGAACCACTCC 1704
535 TyrTyrValLeuGlnIleGluGlnAspAspGlySerGluCysTyrVa 551
1705 TACTACAAAGCTGCAGCTTCTGGAGGAGCACAAGGAAACAGGTATTGGAT 1754
551 lPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLysL 568
1755 ATTCAAGTCTGGGGCGCTGTGGGTAGC...GTATCGGTAGCAACAAAC 1801
568 euGluGluMet...SerLysThrGluAlaIleLysGluPheLysArgLeu 583
1802 TGAACACAGATCCCGTCCAAAGGAGGATGCCATTGAGCAGTTTCATGAATTA 1851
584 PheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAs 600
1852 TATGAAGAAAACCGGGAACGCTTGGCACTCC.....AAAA 1889
600 nPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrGlyV 617
1890 TTTTCAAGAGTATCCCAAAAAGTTTACCCCTGGAGATTGACTATGGCC 1939
617 al...LysLysAlaProLysArgLysAspLysSerGlu...MetLysSer 631
1940 AGGATGAAGAGGAGTGAAGAAGCTCACAGTAAATCTTGGCACCAGTCC 1989
632 SerLeuAlaProGlnLeuLeuGluLeuMetMetLysMetLeuPheAsnValGI 648
1990 AAGCTCCCAAGCCAGTTTCAGGACCTCATCAAGATGATCTTTGATGTGA 2039
648 uThrTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetP 665
2040 AAGTATCAAGAAAGCCATGTGGAGTATGAGATCGACCTTCAGAAGATGC 2089
665 roLeuGlyLysLeuSer*****AsnIleGlu***GlyPheGluAlaLeu 681
2090 CCTTGGGAAGTCAAGCAAAAAGCAGATCCAGCGCCATCTCCATCCCTC 2139
682 Thr*****LeuPheGluGlyHisArg***SerSerThrGlyLe 698
2140 AGT.....GAGGTCCAGCAGCGGCTGTCTCAGGGCGAG 2171
698 u***GluLysAla***LeuLeu*****PheSerLeuLys 715
2172 CAGGACTCTCAGATCTGTGATCTCAATCCCTTTTACACCTGATCC 2221
715 erLeuLeuPheIleLeuIleLeuTyrGlyMetArgMet..... 727
2222 CCACGACTTT.....GGGATGAAGAAGAGCCTCCGCTCCTG 2256
728IleSerTyrSerLysAlaLysMetLeuGluAlaLeuGI 740
2257 AACAAATGCAGACAGTGTGCAGGCCAAGGTGGAATGCTTGACAACTGCT 2306

[illegible][illegible]

PA (ICOS-) ICOS CORP.

XX Christenson E, Demaggio AJ, Goldman PS, McElligott DL;

XX WPI; 2001-025335/03.

DR P-PSDB; AAB47032.

XX New human poly(ADP-ribose) polymerase for treating inflammatory,
PT neurological, cardiovascular, or neoplastic tissue growth disorders,
PT such as, arthritis, encephalitis, myocardial ischemia, and leukocyte
PT metastasis

PS Example 3; Page 116-117; 129pp; English.

XX This sequence encodes the fusion protein PARP1A/PARP2B. This
CC protein contains amino acids 1-662 of hPARP1 fused upstream of
CC amino acids 230-583 of hPARP2. This sequence was amplified using
CC the primer sequences given in AAC85321-40 and AAC85342-51. The fusion
CC protein coding sequence was cloned in a baculovirus expression
CC vector for the expression of the fusion protein. The
CC protein of the invention, hPARP2, causes the covalent addition of
CC polymers of ADP-ribose to protein targets. hPARP2 activity is induced
CC in many instances of oxidative stress or during inflammation where
CC there is direct damage to the DNA. hPARP2 may be used to identify
CC antagonists which may be used to treat a human having a disorder
CC mediated by PARP2 activity, such as, inflammatory, neurological,
CC cardiovascular, or neoplastic tissue growth disorders. hPARP2 and
CC antibodies to it, can also be used to diagnose these conditions.

XX Sequence 3200 BP; 923 A; 745 C; 870 G; 662 T; 0 other;

alignment_scores:

Quality: 1518.50 Length: 1074
Ratio: 2.240 Gaps: 40
Percent Similarity: 63.128 Percent Identity: 35.754

alignment_block:

US-09-236-995d-2 x AAC85341

Align seg 1/1 to: AAC85341 from: 1 to: 3200

6 LysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLysSe 22
127 AAGCTATTCAGTACGATACCCAGAGCGCGCCCTCTTGCAGAA 176
22 rCysArgSerProLeuAlaLysAspGlnLeuArgLeuGlyLysMetValG 39
177 ATGTAGCAGAGCATCCCAAGGACTCGCTCCGGATGGCCATCATGTGTC 226
39 InAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
227 AGTCCCATGTTTGTATGGAAGTCCACACTGTGTACCATCTCTCTGTC 276
56 IlePheSerLysLysAsnGlnIleLysSerValAsp... AspValGluG 71
277 TTCTGGAAGTGGGCCACTCCATCCGACCCCTGAGGTGGGATGG 326
71 yileAspAlaLeuArgTrpAspAspGlnGluLysIleArgAsnTyrVal. 87
327 GTTCTCTAGCTCGGTGGATGACAGAGAGTCAAGAAAGTCAAGACACGG 376
88GlySerAlaSerAlaGluThrSerSerThrAlaAlaProPro 101
377 AAGCTGGAGGAGTACAGGCAAGCCAGGATGGAATTGGTACCAAGGCA 426
102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgth 114
427 GAGAAGACTCTGGGTGACTTTGGCAGCAGAGTATGTCAGTCCACAGAG 476
114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLeus 131
477 TACGTGAAGGGGTGTATGGAGAGATAGAAAAGGGCCAGGTGCGCTGT 526

131 erAlaLysLeu...GluSerGluGlyProLys...GlyIle.....Pro 143
527 CCAAGAGATGGTGGACCCCGGAGGACACACGCTAGCATGATTGACCGC 576
144 TrpTyrHisAlaAsnCysPheGlu.....ValSe 154
577 TGTATACCATCCAGCTGCTTTCAGAACAGGAGGAGGAGCTGGTTTCGG 626
154 rProSerAlaThrValGluLysPheSerGlyTrpAspThrLysSerAspG 171
627 GCCCGAGTACAGTCCGAGTCAGCTCAAGGCTTCAGCCCTCTGTCTACAG 676
171 luAspLysArgThrMetLeuAspLeuValLysLysAsp.....Val 184
677 AGGATAAA.....GAAGCCCTGAAGAGCAGCTCCAGAGATC 714
185 GlyAsnAsnGluGlnAsnLysGlySerLys..... 194
715 AAGAGTGAAGGAAAGAGTAAGCGGATGAGTGGATGGATGATCAAGT 764
195ArgLysLysSerGluAsnAspIleAspSerTyrLysSerAlaArgL 210
765 GGCAGAGAGAAATCTTAAAGAAAGAAAGAC.....AAGGATAGTAAGC 808
210 euAspGluSerThrSerGluGlyThrValArgAsnLysGlyGlnLeuVal 226
809 TTGAAAAAGCC..... 819
227 AspProArgGlySerAsnThrSerSerAlaAspIleGlnLeuLysLeu 243
820CTAAA 824
243 sGluGlnSerAspThrLeuTrpLysLeuLysAspGlyLeuLysThrHisv 260
825 GGTCTAGAACGACCTGATCTGGACATCAAGGACGAGCTAAAGAAAGTGT 874
260 alSerAlaAlaGluLeuArgAspMetLeuGluAlaAsnGlyGlnAspThr 276
875 GTTCAACTAATGACCTGAAGGAGTACTCATCTTCAACAGCAGCAAGTG 924
277 SerGlyProGluArgHisLeuLeuAspArgCysAlaAspGlyMetLeuph 293
925 CCTCTGGGGAGTCGGCGATCTTGGACCGAGTAGCTGATGTCATGCTGT 974
293 eGlyAlaLeuGlyProCysProValCysAlaAsnGlyMetTyrTyrA 310
975 CGGTGCCCTCTCTCCCTCGAGGATGCTCGGTCAGCTGCTCTTCAAGA 1024
310 snGlyGlnTyrGlnCysSerGlyAsnValSerGluTrpSerLysCysThr 326
1025 GCGATGCTATTACTGCACCTGGGAGCTGCTGCTGACCAAGTGTATG 1074
327 TyrSerAlaThrGluProValArgValLysLysLysTrpGlnIleProHi 343
1075 GTCAGACACAGACACCAACCGG.....AAGGAGTGGTAAACCCCA.. 1116
343 sclyThrLysAsnAspTyrLeuMetLysTrpPheLysSerGlnLysVal 360
1117AAGAAATCCGAGAAATCTCTTACCTCAAGAAATTTGAAGTTA 1159
360 ysLysProGluArgValLeuProMetSer..... 370
1160 AAAAGCAGACCGTATATTTCCCAAGAAACCAAGCGCTCCGTGGCGGC 1209
371ProGluLysSerGlySerLysAlaThrGlnArgThrSerle 394
1210 AGCGCTCGCCCTCCACAGCTCGGCTCCCTGCTGCTGTAACCTCTCTGC 1259
394 uLeuSerSerLysGlyLeuAspLysLeuArgPheSerValValGly.... 399
1260 TTCAGCAGATAGGCATATTTCCACATGAAGATCTTCGACTCTCGGGAAGC 1309

```

400 .....GlnSerLysGluAlaAlaAsnGluTrpIleGluLysLeu..... 412
1310 TCTCCCGGAACAAAGGATGAAGTGAAGGCCATGATTGAGAAATCTGGGGG 1359
413 LysLeuAlaGlyAlaAsnPheTyrAlaArgValLysAspIleAspCys 429
1360 AAGTTGACGGGACCGCCACAAAGCTTCCCTG.....TG 1394
429 sLeuIleAlaCysGlyGluLeuAspAsnGluAlaGluValArgLys 446
1395 CATCAGCACCAAAAGAGGTGGAAGAGTGAATGAAGAGTGAAGAGG 1444
446 laArgArgLeuLysIleProIleValArgGluGlyTyrIleGlyLys 462
1445 TAAAGGAAGCAACATCCAGTGTGTCTGAGGAGCTTCTCCAGGAGCTC 1494
463 ValLysArgThrLys.....CysCysHisLeuIleCys 473
1495 TCGGCTCCACCAAGAGCTTCCAGGAGTGTCTTACGCCACATCTGTGTC 1544
473 sIleAsnTrpAsnAlaLeuGluSerLysGly***ThrValThrVal 490
1545 C...CCTTGGGGGACAGAGTGAAGGAGAGCTGTGTAAGTT...GTGG 1588
490 ysValLysGlyArgSerAlaCysSer***SerPro***ValCysLysAsn 506
1589 CCCAAGAGGAGTCCAGGGCTGCCTCTCCAAA.....AAGAGC 1629
507 ThrAlaHisIlePro***Trp.....GluLysHisIle 519
1630 AAGGGCCAGGTCAAGGAGGAGGTATCAACAATCTGAAGAGAGATGAA 1679
519 nCys***LeuLys..... 523
1680 ATTAAGTCTTAAAGAGGAGGAGAGCTGTGGATCCTGATCTGACAGTGAAC 1729
524 .....HisValLeu.....Thr*** 528
1730 ACTCTGGGATCTCTGGAGAAAGTGGGAGGTCTTCAAGTCCACCCCTT 1779
529 His***Val.....CysThrGlyTyrTyrValLeuGlnIle 540
1780 GGCCTGTGGACATCGTTAAAGGAACCACTCTACTACAAAGTGCAGCT 1829
540 eIleGluAlaAspAspCysSerGluCysTyrValPheArgLysTrpGly 557
1830 TCTGGAGGACGAAGAAACAGGTATTGGATATTACAGTCTCTGGGGC 1879
557 rgValGlySerGluLysIleGlyGlnLysLeuGluMet...Ser 572
1880 GTGTGGGTAG...GTGATCGGTAGCAACAACTGGAACAGATGCCGTCC 1926
573 LysThrGluAlaIleLysGluPheLysArgLeuPheLeuGluLysThrG 589
1927 AAGGAGGATGCCATTTGACACTTTCATGAAATATATGAAGAAAAACCG 1976
589 yAsnSerTrpGluAlaTrpGluCysLysThrAsnPheArgLysGlnPro 606
1977 GAAGCTTGGACATCC.....AAAATTTCCAGCAAGTATCCCA 2014
606 lyArgPheTyrProLeuAspValAspTyrGly.....Val 617
2015 AAAAGTTCTACCCCTGGAGATTGACATATGCCAGGATGAAGAGGACGT 2064
618 LysLysAla.....ProLysArgLysAspIleSerGluMetLys 631
2065 AAGAAAGTGCAGTAATCTCTGGACCAACAGCGTTCACAGATCACAGCT 2114
631 rSerLeuAlaProGlnLeuLeuGluLeuMetLysMetLeuPheAsnVal 648
2115 AGATCTTGGGTACAG.....GAGTTAATAAAGTTGATCTGTAATGTC 2158
648 luThrTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMet 664

```

3026 TAGGACGACGAAGTGACACAGGAATTCGAATCCAGATGTTATACCCCTC 3075
 956 MetTyrAsnGluTyrIleValTyrAsnThrSerClnValLysMetGlnPh 972
 3076 AACTACAATGAATATATTGTATATAACCCCAACCCAGGTCCGTATGCGGTA 3125
 972 eLeuLeuLysValArgPheHis 979
 3126 CCTTTAAAGGTTTCAGTTTAA 3147

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA1995.DAT: AAT04221

seq_documentation_block:

ID AAT04221 standard; DNA; 5345 BP.

AC AAT04221;

DT 31-MAR-1996 (first entry)

DE 5.3 kb EcoRI fragment containing poly(ADP-ribose)-polymerase gene.

XX DNA primer; PCR; polymerase chain reaction;

KW poly(ADP-ribose) polymerase; cancer diagnosis; chromosome-13;

KW pseudogene; Burkitt's lymphoma; B-lymphoma; colorectal carcinoma; tumor;

KW small cell lung carcinoma; colorectal carcinoma; tumor;

KW breast carcinoma.

XX Homo sapiens.

OS

XX US5449605-A.

PN

XX 12-SEP-1995.

XX 14-OCT-1988; 88US-0257696.

XX 06-APR-1993; 93US-0044618.

PR 14-OCT-1988; 88US-0257696.

XX (GEOU) UNIV GEORGETOWN.

XX Cherney B, Lyn D, Smulson ME;

PI WPI; 1995-327692/42.

XX Allele-specific primers for detecting pre-disposition to cancer -

PT caused by a deletion in chromosome 13 pseudo-gene for poly

PT (ADP-ribose) polymerase

XX Disclosure; Page 41-46; 47pp; English.

XX This sequence corresponds to a 5.3 kb EcoRI fragment that contains

CC the poly(ADP-ribose)-polymerase pseudogene.

XX Sequence 5345 BP; 1598 A; 1142 C; 1396 G; 1209 T; 0 other;

SQ

alignment_scores:

Quality: 1237.00 Length: 1128

Ratio: 1.811 Gaps: 49

Percent Similarity: 60.550 Percent Identity: 33.511

alignment_block:

US-09-236-995D-2 x AAT04221 ..

Align seg 1/1 to: AAT04221 from: 1 to: 5345

6 LysAlaTrpLysAlaGluTyrAlaLysSerGlyArgAlaSerCysLysSe 22

791 AAGCTCTACTGAGTGAGTACACCAAGAGCGGCGCACCTCTTCAAGAA 840

22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetValG 39

841 ATGCAGCGAGAGCATCCCAAGGAGTCCGTCGATGCGCCATCATGTGTC 890

39 InAlaSerGlnPheAspGlyPheMetProMetTyrAsnHisAlaArgCys 55
 891 AATAGTCCATGTT.GATGGAGAAGTCCCACACTGGTACCACCTTCTCTCTC 939
 56 IlePheSerLysLysAsnGlnIleLysSerValAsp...AspValGluG1 71
 940 TTGCAGAGAGTGGCGCACCTCCATGGCGCACCTGACGTTGAGTGGATGG 989
 71 yIleAspAlaLeuArgTyrAspGlnGluLysIle.....ArgA 85
 990 ATTCTCTGAGCTTTGGTGGGATGACACGAGAAAGTCAAGAGACACGTG 1039
 85 snTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
 1040 ACGCTGGAGGAGTGACGGGCAAAAGCCAGATGGAATTGGTAGCAAGGCA 1089
 102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
 1090 GAGAAGACGCTGGGTGACTTTGCAGCAGAGTAGGCCAAGTCCAGCAGAAG 1139
 114 rSerCysArgArgCys.SerGluLysIleThrLysGlySerValArgLeu 130
 1140 CAAGTGCAGAGAGTGTATGGAGAGATAGAAAAGGCCAGATGCGCTA 1189
 131 SerAlaLys....LeuGluSerGluGlyProLys...GlyIle.....Pr 143
 1190 TCCAAGAAGATGCTGGACCCCTGAGAAGCCTCAGTAGGCATGATTGACOC 1239
 143 oTrpTyrHisAlaAsnCysPhePheGlu.....Vals 154
 1240 CTGGTACCACCCAGGCTGCTTGTCAAGAACAGAGGAGGAGTGGGTTCOC 1289
 154 erProSerAlaThrValGluLysPheSerGlyTyrAspThrLeuSerAsp 170
 1290 AGCTGAGTACAGTGGAGTCAAGGCTTCAGCCTCTCTCGCTGCA 1339
 171 GluAspLysArgThrMetLeuAspLeuValLysLysAspValGlyAsnAs 187
 1340 GAGGATAAAGAACCCCTG.....AAGAAGCAGCTCCAGGAGT 1377
 187 nGluGlnAsnLysGlySerLysArgLysLysSerGluAsnAspIleAsp 204
 1378 CAAG.....AGTGAAGAAAGAGAGAACGCGATGAGGTGGATG 1415
 204 erTyrLysSerAlaArgLeuAspGluSerThrSerGluGlyThrValArg 220
 1416 GA.....GTGGATGA.....GTGGCCAAG 1435
 221 AsnLysGlyGlnLeuValAspProArgGly.SerAsnThrSerSerAlaA 237
 1436 AAGAAATCTAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1484
 237 spIleGlnLysLysLysGluLysSerAspThrLeuTrpLysLysLys 253
 1485CTAAGGCCCAAGACGCTGATCTGGAAACATCAAG 1520
 254 AspGlyLeuLysThrHisValSerAlaAlaGluLeuArgAspMetLeuG1 270
 1521 GACGAGCTAATGAAGTGTCTTATTAAAGCTGACCTGAGGAGATGCTCAT 1570
 270 uAlaAsnGlyGlnAspThrSerGlyProGluArgHisLeuLeuAspArgC 287
 1571 CTTCAACAGGACGAGTGCCTTCGGGGAGTGGCGGATCTTGGACCGAG 1620
 287 ysAlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAla 303
 1621 TAGCTGACAGCATGCTGGTGGCTCTCTCCCTGAGGAGATGCTCA 1670
 304 AsnGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerGlyAsnValSe 320
 1671 GGTGAGTGTCTTCAAGAGCGACACTTATTACTGCACCGGGGACGCTCAC 1720

480 TCCTTACTCTTGGGAAGTCTCCCGGAACAAGGATGAAGTGAAGGCCACG 529
409 IleGluLysLeuLysLeuAlaGlyAlaAsnPheTyrAlaArgValVal 425
|||||
530 ATTGAGAACTC.....GGGGAAAGTTGATGGGACGGCCAAACA 570
425 sAspIleAspCysLeuIleAlaCysGlyGluLeuAspAsnGluAsnAlaG 442
|||||
571 GGCTCTCTCGTCATCAGCACTAAAGAGAGGGGTGGAAGAAAGAAAGA 620
442 luValArgLysAlaArgArgLeuLysIleProIleValArgGluGlyTyr 458
|||||
621 AGATGAGGAAGTGAAGAAAGCAACACTGAGCTGTCTGAGGACTTC 670
459IleGlyGlu.CysValLysA 465
671 CTCAGGACTTCTCTGCCCTCCACCAAGGGCTCCAGGAGTTGTCTCAGC 720
465 rgThr..... 466
721 GCACATATTGACCCCTTGGGGGCGACAGGTGAAGGCAGAGCCTGTGTAAG 770
467LysCysCysHis..... 470
771 TCGTAGCCCCAAGAGGAAGTCAAGAGCTGTCTCTCCAAAAAAGCAAG 820
470 470
821 GGCCAGGTCAAGGAGGAGGTATCAACAAATCTGAAAGAGAATGAAATT 870
471LeuIleCysIleAsnTrpAsnAla 478
871 AACTCTTAAGGAGGAGCAGCTGTGGATCCTGACTCTGCTGTGGAACACT 920
479 LeuGluSerSer..... 482
921 CTGCGCATCTCTGGAGAAAGGTGAAGGCAGAGCCTGTTGAAGTCTGAGC 970
483LysGly***ThrValThrValLysValLysGlyArg. 494
971 CCAAGAGGGAAGTCAAGAGCTGTGCTCTCCAAAAAAGCAAGGCCACG 1020
494 494
1021 TCAAGGAGGAGGTATCAACAACATCAAAAGAGATGAATTAACCTTT 1070
495SerAlaCysSer***SerPro***ValCysLysAsnThrAlaHi 509
1071 AAAGGAGGAGCAGCTCTGGATCTGACTCTGCTGCTGGAACACTCTGCACA 1120
509 sIlePro***TrpGluLysHisIleGlnCys***LeuLysHisValL 526
1121 TGTTCTGGAGAAAGTGGGAAGGTCTTCAAGTCCACCTCAGCCCTGCTGGT. 1169
526 euThr***His***ValCysThrGly.....TyrTyrValLeuGln 539
1170GACGTCTGTAAGGAACCACTCTCTATTACAAGTGAAGT 1208
540 IleIleGluClnAspAspGlySerGluCysTyrValPheArgLysTrpGl 556
1209 TTGCTGAAGGATGACAAGAAAGCAGGCATTTGATATTCAAGTCTCTGGA 1258
556 YArgValGlySerGluLysIleGlyGlnLysLeuGluMet...S 572
1259 CCGTGTGGGCACG...GTGATCGGTAGCAACAACCTGGACAGAGTCTCT 1305
572 erLysThrGluAlaIleLysGluPheLysArgLeuPheLeuLysThr 588
1306 CCAAGGAGGACACCATTAACACTTTCATGAAATATTATGAAGAAAACTA 1355
589 GlyAsnSerTrpGluAlaTrpGluCysLysThrAsnPheArgLysClnPr 605
1356 GGAAT.....GCTTGGCACTCCAAAA...TTCAAAAGTATCC 1391

605 oGlyArgPheTyrProLeuAspValAspTyrGlyVal...LysLysAlaP 621
1392 CAAAAGTTCTACCCCTGGAGATTGACTACGCCAGGACAAAGAGCGG 1441
621 roLysArgLysAspIleSerGluMetLysSerSerLeuAlaProGln... 636
1442 TGAAGAGCCGACAGTAATCTCTGGCACCACCAAGTCCATGCTCCCAAGCCA 1491
637 LeuLeuGluLeuMetLysMetLeuPheAsnValGluThrTyrArgAlaAl 653
1492 GTTCAGGACCTATCAAGAATGATCTTTGATGTGAAATATGCAAGGCC 1541
653 aMetMetGluPheGlu***AsnMetSerGluMetProLeuGlyLysLeus 670
1542 CCATGGGGGTGTGAGATCAACTT...CAGATGCCCTTGGGAAGCTGA 1588
670 er****AsnIleGlu***GlyPheGluAlaLeuThr***** 686
1589 GCAAAAGGCAATCCAGGCCGCTACTCCATCTCAGGTC..... 1628
687 LeuPheGluGlyHisArg***SerSerThrGlyLeu***GluLysAla** 703
1629CAGCAGGTGTGTCTCCAGGCGCAGCAGCACTCTCAGAT 1666
703 *LeuLeu*****PheSerLeuLeuSerLeuPheIleL 720
1667 CCTG.....GATCTCTCAATCGCTTTTACATCC 1695
720 euIleLeu.....TyrGlyMetArg.....MetIleSer..... 729
1696 TGATCCCCACGAGCTTTGGGATGAAGGATCCTCTGCTCTCTGAAACAATGCA 1745
730TyrSerLysAlaLysMetLeuGluAlaLeuGlnAspIleGl 743
1746 GACAGTGTGAGGCCAAGGTAGAAATGCTGCACAACTCTGTGACATGTA 1795
743 uIleAlaSerLysIleVal.....GlyPheAspSerAspSerAspGluS 758
1796 GGTAGCTACGGTCTGCTCAGGGGAGGCTCTCAGCATAGCAGGAAGGACT 1845
758 erLeuAspAspLysTyrMetLysLeuHisCysAspIleThrProLeuAla 774
1846 CCATCGATGTCAACTATGAGAAGCTCAAACTGACATTAAGTGGTGGTGC 1895
775 HisAspSerGluAspTyrLysLeuIleGluGlnTyrLeuLeuAsnThrHi 791
1896 AGATTTCTGAAGAGCTGAGATCATCAGGAGTATGTTAAGACACTCA 1945
791 sAla.Pro...ThrHisLysAspTrpSerLeuGluLeuGluValPhe 806
1946 TGCAACCAACACACACGATCATATGACTTGAAGTCAATTGATAGCTTT 1995
807 SerLeuAspArgAspGlyGluLeuAsnLysTyrSerArgTyrLysAsnAs 823
1996 AAGATAGAGTGTGAAGAGGAGTGCACACCTACAAAGCCCTTTAAG...CA 2042
823 nLeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheV 840
2043 GCTTCATACCTGAAGTGTCTGTGGCATGCTCCAGGACCACTTGG 2092
840 alGlyIleLeuSerGlnGlyLeuArgIleAlaProProGluAlaProVal 856
2093 CTGGGATCTCTCCTGGTCTTTGGATAGCCCTGCCTGAAGCACTGTG 2142
857 ThrGlyTyrMetPheGlyLysGlyLeuPheAlaAspLeuValSerLy 873
2143 ATGGGCTACATGTTGGTAAAGTAGTCTATTTCGTCATCTTGCTCTCAA 2192
873 sSerAlaGlnTyrCystyrValAspArgAsnAsnProValGlyLeuMetL 890
2193 GAGTGCCAAACGACTGCCATACATCTTAGAAGACCAATAGGTTAATCC 2242

254 pGlyLeuLysThrHisValSerAlaAlaGluLeuArgAspMetLeuGluA 271
 858 CGAGCTAAAGAAAGTGTGTTCAACTAATACCTGAAGGAGCTACTCATCT 907
 271 laAsnGlyGlnAspThrSerGlyProGluArgHisLeuLeuAspArgCys 287
 908 TCAACAAGCAGCAAGTGTCTCTGGGAGTCTGGCGATCTTGGACCGAGTA 957
 288 AlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAlaAs 304
 958 GCTGATGGCATGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1007
 304 nGlyMetTyrTyrTyrAsnGlyGlnTyrGlnCysSerGlyAsnValSerG 321
 1008 TCAGCTGGTCTTCAAGAGCGATGCTTACTGCACTGGGGAGCTCACTG 1057
 321 luTrpSerLysCysThrTyrSerAlaThrGluProValArgValLysLys 337
 1058 CCTGGACCAAGTGTATGGTCAAGACACAGACACACCCCAACCGG....AAG 1101
 338 LysTrpGlnTleProHisGlyThrLysAsnAspTyrLeuMetLysTrpPh 354
 1102 GAGTGGTAAACCA....AGGAATCCGAGAAATCTCTTACCT 1142
 354 eLysSerGlnLysValLysLysProGluArgValLeuProMetSer. 370
 1143 CAAGAAATTCAGGTTAAAGAGCAGGACCGTATATTCCTCCCGCAGAAACCA 1192
 371ProGluLysSerGlySerLysAla 378
 1193 GCGCCTCCGTGGGGGCGCCCTCCGCGCTCCACAGCCTCGGCTCTGCT 1242
 379 ThrGlnArgThrSerLeuLeuSerSerLysGlyLeuAspLysLeuArgPh 395
 1243 GCTGTGAACCTCTCTGCTTCCAGCAGATAGCCATTTATCAACATGAAGAT 1292
 395 eSerValValGly.....GlnSerLysGluAlaAlaAsnGluTrpI 409
 1293 CCGTACTCTCGGAAGCTGTCCCGGCAACAGGATGAAGTGAAGGCCATGA 1342
 409 leGluLysLeu.....LysLeuAlaGlyAlaAsnPheTyrAlaArgVal 423
 1343 TTGAGAACTCGGGGGAAGTTGACGGGGAGCGCCCAACAGGCTTCCCTG 1392
 424 ValLysAspLysLeuLeuAlaCysGlyGluLeuAspAsnGluAs 440
 1393TGCATCAGCACCAACCAAGAGGAGTGAAGGATGAA 1427
 440 nAlaGluValArgLysAlaArgLeuLysIleProIleValArgGluG 457
 1428 TAAGAAATGGAGGAAGTAAAGGAAGCAACATCCGAGTTGTGTGAGG 1477
 457 lyTyrIleGlyGluCysValLysArgThrLys..... 467
 1478 ACTTCTCCAGGAGCTCTCCGCTCCACCAAGAGCCTTCAGGAGTTGTT 1527
 468 CysCysHisLeuIleCysIleAsnTrpAsnAlaLeuGluSerSerLysG 484
 1528 TTAGCGCACATCTTGTGCC....CCTTGGGGGAGAGAGTGAAGGCAGAGCC 1574
 484 y**ThrValThrValLysValLysGlyArgSerAlaCysSer***SerP 501
 1575 TGTGGAAGTT...GTGGCCCCCAAGGGAAGTCAAGGGGCTCGGCTCCTCCA 1621
 501 ro***ValCysLysAsnThrAlaHisIlePro*****Trp..... 514
 1622 AA.....AAAGCAAGGCCAGGTCAAGGAGGAAGGTATCAACAAA 1662
 515 ...GluLysHisIleGlnCys***LeuLys..... 523
 1663 TCTGAAAGAGAAATTAACCTTAAAGGAGGAGGACGCTGTGGATCC 1712
 524HisValLeu..... 526
 1713 TGATCTGGAGTGAACACACTCTCGCATGCTCTGGACAAAGGTGGGAAG 1762
 527Thr***His***Val.....CysThrGly 534
 1763 TCITTCAGTGCACCCCTTGGCTGTGTGCATCGTTTAAAGGAACCAACTCC 1812
 535 TyrTyrValLeuGlnIleIleGluGlnAspAspGlySerGluCysTyrVa 551
 1813 TACTACAGCTGTCAGCTTCTGGAGGACGACAGAAACACAGGATTGGAT 1862
 551 lPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLysL 568
 1863 ATTCAAGTCTGGGGCGCTGTGGTAGS...GTGATCGGTAGCAACAAC 1909
 568 euGluGluMet....SerLysThrGluAlaIleLysGluPheLysArgLeu 583
 1910 TGAACAGATGCCGTCCAAAGGAGGATGCCATGTGACACTTTCATGAATA 1959
 584 PheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAs 600
 1960 TATGAAGAAACACCGGAGACCTTGGCACCTCC.....AAAA 1997
 600 nPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrGly. 616
 1998 TTTCACGAAGTATCCCAAAAGTTCTACCCCTCGGAGATTGACTATGCC 2047
 617ValLysLysAla.....ProLysArgLysAsp 625
 2048 AGGATGAAGAGGCGAGTGAAGAAGCTGACAGTAAATCTTGGCACCACAACGC 2097
 626 lle..... 626
 2098 GTTGAGACAAACACCTTTAGATTAGTTTTCAGCAGATGTCAGCGC 2147
 627SerGluMetLysSerSerLeuAlaP 635
 2148 TCCTCTGACAGCAGCATGCCCTCTGCTCTGCTCTGTTTACAAAGC 2197
 635 roGlnLeuLeuGluLeuMetLysMetLeuPheAsnValGluThrTyrArg 651
 2198 CTCAAAGTGTCAATGGTGTGAGA.....AGCCAGGAGCCACTGCA 2238
 652 AlaAlaMetMetGluPheGlu***AsnMetSerGluMetProLeuGlyLy 668
 2239 GATGCTCTCTTCAGGTCATCTAGCCCATCAAGCCTTCTGACAGCCAG 2288
 668 sLysSer*****AsnIleGlu***GlyPheGluAlaLeuThr***** 685
 2289 C...AGTCTTGACAACTTATCTGGGAGTTTTCAGAACTGCTTCCAGTA. 2334
 685 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu***GluLys 701
 2335GTTAGTTCAAGTGA...ACAGAGGT 2358
 702 Ala***LeuLeu*****pHeSerLeuLe 714
 2359 GCTTCCAGTTTGGAGAAAAAGGAGGTTCCAGGAGTAGATTTTAGCATAAC 2408
 714 uSerLeuLeuPheIleLeuIleLeuTyrGlyMetArgMetIleSerTyrS 731
 2409 TCAATTCTGAAGAAATCTTGGAGCTTGGACCTTATGGATATATTTGACA 2458
 731 erLysAlaLysMetLeuGluAlaLeuGlnAspIle..... 742
 2459 GAGACAGATCACTTTGGATGTTAGTTGAGATGGGCAAGGAGCTG 2508
 743 ...GluIleAlaSerLysIleValGlyPheAspSerAspSerAspGluSe 758
 2509 AAGGAGATTGGAATCAATCTTATGACATAGGACAACTAATTAAGG 2558
 758 rLeuAsp.....AspLysTyrMetLysL 766

```

11-MAY-2001 (first entry)
Human poly(ADP-ribose) synthetase sbhPARS2 cDNA.
Human: poly(ADP-ribose) synthetase; sbhPARS2; vaccine; drug screening;
ischaemic disorder; cerebral ischaemia; cardiac ischaemia;
myocardial infarction; stroke; inflammation; autoimmune disease;
diabetes; multiple sclerosis; neurodegenerative disease;
Parkinson's disease; Alzheimer's disease; chromosome localisation; ss.
Homo sapiens.
Key Location/Qualifiers
CDS 1..1566
FT /*tag= a
PE /product= "Human poly(ADP-ribose) synthetase sbhPARS2"
XX WO200112645-A1.
XX 22-FEB-2001.
XX 10-AUG-2000; 2000WO-US21775.
XX 12-AUG-1999; 99US-0373441.
XX (SMK ) SMITHKLINE BEECHAM CORP.
XX Barone F, Field J, Kabnick KS, Li X, McQueney MS, Zhu Y;
XX WPI; 2001-211196/21.
XX P-PSDB; AAB60693.
XX New human polypeptide of the polyADPribose synthetase family for use in
XX diagnostic assays and for screening modulators used for preventing and
XX treating inflammation, autoimmune disease and Alzheimers disease
XX Claim 2; Page 21; 30pp; English.
XX This sequence represents cDNA encoding a novel human poly(ADP-ribose)
XX synthetase, sbhPARS2. The invention also relates to fragments,
XX variants and sequences with at least 95% identity to the sbhPARS2
XX protein or nucleotide sequence; expression systems and host cells
XX comprising an sbhPARS2 nucleic acid sequence; the recombinant expression
XX of sbhPARS2; and an antibody specific for sbhPARS2. sbhPARS2 proteins
XX and nucleotides are useful as vaccines for inducing an immunological
XX response in a mammal. The sbhPARS2 protein is useful for identifying
XX compounds which inhibit or stimulate its activity or expression level.
XX Such agonists and antagonists of sbhPARS2 are useful for treating human
XX diseases including ischaemia and ischaemic tissue injury (e.g., cerebral
XX and cardiac ischaemia, myocardial infarction, stroke), inflammation,
XX autoimmune disease (e.g. diabetes, multiple sclerosis) and
XX neurodegenerative diseases (e.g., Parkinson's disease and Alzheimer's
XX disease). sbhPARS2 nucleic acids are useful as diagnostic reagents for
XX detecting mutations in the associated gene; as hybridisation probes
XX to isolate full-length sbhPARS2 cDNAs and sbhPARS2 genomic clones; and
XX as an immunogen to produce antibodies for therapeutic use. sbhPARS2
XX proteins, nucleotides and antibodies are also useful in screening methods
XX for detecting the effect of added compounds on the production of mRNA and
XX protein in cells.
XX Sequence 1566 BP; 490 A; 328 C; 381 G; 367 T; 0 other;

```

```

alignment_scores:
  Quality: 851.50      Length: 499
  Ratio: 2.534        Gaps: 15
Percent Similarity: 67.335      Percent Identity: 40.882
alignment_block:
US-09-236-995D-2 x AAF59996

```

seq_name: /SDS2/gc9data/geneseq/geneseq/NA2001.DAT:AAF59996

seq_documentation_block:

ID AAF59996 standard; cDNA; 1566 BP.

XX

AC AAF59996;

Align seg 1/1 to: AAF59996 from: 1 to: 1566

```
519 GlnCys***LeuLysHisValLeuThr***His***ValCysThrGly... 534
:::||||| ||| ||| ||||| ||| ||| |||
103 GAGTGTACAGCAAG...GTGGGAGGCTCATGTGTATTGTGAAGGAA 149
534 .....
150 TGATGCTATGATGTCATGCTAAATCAGACCAATCTCCAGTTCAACAACA 199
535 .....TyrTyrValLeuGlnIlelleGluGlnAspAspGlySerGluCys 549
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 ACAAGTACTATCTGATCTAGTATAGAGATGATGATGATGATGATGATG 249
550 TyrValPheAlaGlyLysTyrGlyArgValGlySerGluLysIleGlyGly 566
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 AGTGTGTTGGATGAGATGGGGCGAGTTGGG.....AAATGGGACAGCA 293
566 nLysLeuGluGluMetSer.....LysThrGluAlaIleLysGluPheL 581
:::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
294 CAGCCTGTGTGCTTCTCAGGCAATCTCAACAAGGCCCAAGGAATCTTTC 343
581 ysArgLeuPheLeuGluLysThrGlyAsnSerTyrGluAlaTyrPheGlyCys 597
:::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 AGAAGAAATCTCTTGACAAAAACGAAAAACAATTTGGGAAGATCGAGAA... 390
598 LysThrAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValAs 614
:::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 .....AAGTTTGAAGAGTGCCTGGAANAATATGATGTATGATGATGGA 629
614 pTyrGlyValLysLysAlaProLysArgLysAspIleSerGlu.....M 629
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
435 CTATCCCACTACTCTCAGGATGAGAGGAAACAAAGAAAGAGGAAATCTC 484
629 eLysSerSerLeuAlaProGln.....LeuLeuGlu 639
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
485 TTAATATCTCCCTTGAAGCCAGAGTACAGCTAGATCTTCGGGTACAGAG 534
640 LeuMetLysMetPheAsnValGluThrTyrArgAlaAlaMetMetG 656
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
535 TTAATATAAGTTGATCTGTAATGTTGAGGCCATGGAAGAAATGATGATGGA 584
656 pHeGlu***AsnMetSerGluMetProLysLysLysLeuSer*****A 673
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
585 AATGAAGTATAATACCAAGAAAGCCCACTTGGGAAGCTGACAGTGGCAC 634
673 snIleGlu***GlyPheGluAlaLeuThr*****LeuPheGlu 689
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
635 AAATCAAGGCAGGTTACCAGTCTCTTAAGAAGATTGAGGATTGTATTGG 684
690 GlyHisArg***SerSerThrGlyLeu***GluLysAla***LeuLeu** 706
:::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
695 .....CCTGGCCAGCATGGACGAGCTCTCATGGAAGC 716
706 *****PheSerLeuLeuSerLeuLeuPheIleLeu..... 720
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
717 ATGCAATGAATTTCTACACAGGATTCGCGATGACTTTGGACTCCGTACTC 766
721 .....IleLeuTyrGlyMetArgMetIleSerTyrSerLysAlaLysMet 735
767 CTCCTACTAATCCGGACACAGAGGAAGTGTCA...GAAAAAATACAAATTA 813
736 LeuGluAlaLeuGlnAspIleAlaSerLysIleValGlyPheAs 752
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
814 CTAGAGGCTTTGGGAGACATTTGAATGCTATTAGCTGGTGGTGAACAGCA 863
752 pSerAspSerAspGlu...SerLeuAspAspLysTyrMetLysLeuHisC 768
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
864 GCTACAAAGCCCAAGAACCCCACTTGGACCACTATAGAAACCTACATT 913
768 ysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeuIleGlu 784
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
914 GTGCCTTGGCCGCCCTTGACCATGAAGATTATGAGTTCAAAAGTATTTC 963
```

```
785 GlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTyrSerLeuG 801
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
964 CAGTACCTACAAATCTACCATGCTCCACACACAGCGACTATACCATGAC 1013
801 uLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTyrS 818
||| ::::::::::::::::::::|||
1014 CTTGCTGGATTTGTTTGAAGTGAGAGGATGGTGAG.....AAAG 1054
818 erArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTyrPheHisGlySer 834
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 AACCTTTCAGAGGAGACCTTCATACAGAGATGCTCTATGCGCATGGTTC 1104
835 ArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAlaPr 851
||| ::::::::::::::::::::|||
1105 AGGATGAGTAAGTGGCTGGGAATCTTTGAGCCATGGGCTTCGAATTCGCC 1154
851 oProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheA 868
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 ACCTGAAGCTCCCATCAGAGTTACATGTTGGGAAAGGAATCTACTTTG 1204
868 laAspLeuValSerLysSerAlaGlnTyrCystTyrValAspArgAsnAsn 884
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1205 CTACATGCTCTCCAAGAGTGCCAAATCTGCTTTGCCCTCTGCCCTAAAG 1254
885 ProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMetTyrG 901
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1255 AATACAGGACTGCTGCTCTATCAGAGGTAGCTTAGGTGAGTGAATGA 1304
901 uLeuLysLysAlaThr...SerMetAspLysProProArgGlyLysHisS 917
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1305 ACTACTAGAGGCCAATCTTAAGGCCGAAGGATTGCTTCAAGGTAACATA 1354
917 erThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheValLys 933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1355 GCACCAAGGGCTGGCAAGATGGCTCCCGATGCTGCCCATCTTCGTCACC 1404
934 TrpArgAspAspValValValProCysGlyLysProValProSerSerI 950
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405 CTGAATGGGAGT...ACAGTGCCATTAGGACCCAGCAAGTGACACAGGA 1451
950 eArgSerSerGlu.....LeuMetTyrAsnGluTyrIleValTyrA 964
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1452 TCTGAATCCAGATGGTTATACCTCACTCACTAATGAATATATGTATATA 1501
964 snThrSerGlnValLysMetGlnPheLeuLysValArgPheHis 979
||| ::::::::::::::::::::|||
1502 ACCCCACAGGTCCTGATCGGTACCTTTTAAAGGTTTCAGTTTAAAT 1548
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAC85303

seq_documentation_block:

ID AAC85303 standard; cDNA; 1814 BP.

XX AAC85303;

XX 29-MAR-2001 (first entry)

XX hparp2 cDNA.

DE Human; poly(ADP-ribose) polymerase; hPARP2; oxidative stress; ARDS;
KW inflammation; ischaemic stroke; hemorrhagic shock; myocardial ischemia;
KW infarction; cerebral vasospasm; rheumatoid arthritis; osteoarthritis;
KW gouty arthritis; spondylitis; Behcet's disease; sepsis; septic shock;
KW endotoxic shock; gram negative sepsis; gram positive sepsis; trauma;
KW toxic shock syndrome; multiple organ injury syndrome; vasculitis;
KW hemorrhage; conjunctivitis; uveitis; thyroid-associated ophthalmopathy;
KW eosinophilic granuloma; asthma; chronic bronchitis; allergic rhinitis;
KW chronic obstructive pulmonary disease; silicosis; reperfusion injury;
KW pulmonary sarcoidosis; pleurisy; alveolitis; pneumonia; myocardium;
KW bronchiectasis; pulmonary oxygen toxicity; keloid formation; brain;
KW scar tissue formation; atherosclerosis; systemic lupus erythematosus;
KW autoimmune thyroiditis; multiple sclerosis; Reynaud's syndrome;

736 LeuGluAlaLeuGlnAspIleGluLeuAlaSerLysLysIleValGlyPheAs 752
|||||
1062 CTAGAGGCTTTGGGAGACATTGAATGCTATTAGCTGGTGAACACAGA 1111
752 pSerAspSerAspGlu...SerLeuAspAspLysTyrMetLysLeuHisC 768
FT : : : : :
1112 GCTACAAAGCCAGAACACCCATTGGACACACACTATAGAAACCTACATT 1161
FT : : : : :
768 ysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeuIleGlu 784
PN : : : : :
1162 GTGCCTTGGCCCTTGCATGCAATGTTACGAGTTCAAAAGTATTTC 1211
FT : : : : :
785 GlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSerLeuG1 801
XX : : : : :
1212 CAGTACCTACAATCTACCATGCTCCACACACAGCAGCTATACCATGAC 1261
FT : : : : :
801 uLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTyrS 818
XX : : : : :
1262 CTTCGCTGGATTGTTGAAGTGGAGAGGATGGTGAG.....AAAG 1302
FT : : : : :
818 erArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlySer 834
XX : : : : :
1303 AAGCCTTCAGAGAGGACCTTCATAACAGAGTCTTCTATGCGATGGTTCC 1352
FT : : : : :
835 ArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAlaPr 851
XX : : : : :
1353 AGGATGAGTAAGTGGTGGGAATCTTGAGCCATGGCTTGAATGGCCCC 1402
FT : : : : :
851 oProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheA 868
XX : : : : :
1403 ACCTGAAGCTCCCATCACAGGTTACATGTTTGGGAAGGAATCTACTTTG 1452
FT : : : : :
868 laAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArgAsnAsn 884
XX : : : : :
1453 CTGACATGCTCTTCCAAGAGTGGCAATTAAGTCTTGGCTCTCGCCTAAAG 1502
FT : : : : :
885 ProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyrG1 901
XX : : : : :
1503 AATACAGGACTGCTGCTCTTATCAGAGGTAGCTCTAGGTCAGTGAATGA 1552
FT : : : : :
901 uLeuLysLysAlaThr...SerMetAspLysProProArgGlyLysHisS 917
XX : : : : :
1553 ACTACTAGAGGCGCAATCTTAAGCGGAGGATTTGCTTCAAGGTAACATA 1602
FT : : : : :
917 erThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheValLys 933
XX : : : : :
1603 GCACCAAGGGGCTGGGCAAGATGGCTCCCGATTCGCCACTCGTCACC 1652
FT : : : : :
934 TrpArgAspValValProCysGlyLysProValProSerSerI1 950
XX : : : : :
1653 CTGAATGGGAGT...ACAGTGCCATTAGGACCCAGCAAGTGACACAGGAAT 1699
FT : : : : :
950 eArgSerSerGlu.....LeuMetTyrAsnGluTyrIleValTyrA 964
XX : : : : :
1700 TCTGAATCCAGATGGTTATACCTCACTCACTACATGAATATATGTTATATA 1749
FT : : : : :
964 snThrSerGlnLysMetGlnPheLeuLeuLysValArgPheHis 979
XX : : : : :
1750 ACCCAACACAGGTCGCTATGCGGTACCTTTAAAGTTTCAGTTTAAAT 1796
FT : : : : :
seq_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ44287

ischemic tissue damage; ss.
Homo sapiens.
Key Location/Qualifiers
CDS 3..1715
/*tag= a
/product= "PARP2"
WO9964572-A2.
16-DEC-1999.
04-JUN-1999; 99WO-EP03889.
05-JUN-1998; 98DE-1025213.
01-MAR-1999; 99DE-1008837.
(BADI) BASF AG.
Kock M, Hoeger T, Kroeger B, Otterbach B, Lubisch W, Lemaire H;
WPI: 2000-087218/07.
P-PSDB; AAY51174.
Novel genes and proteins, antibodies and binding partners useful in
diagnosis and therapy of energy deficiency associated disease
conditions -
Claim 7a; Page 49-52; 96pp; German.
This invention describes novel human and murine poly(ADP-ribose)
polymerase (PARP) homologues, which are characterised by an amino acid
sequence with a functional NAD⁺-binding site and no zinc finger
sequence motif, of general formula CX₂CX₂MX₂2C (I). The nucleic acid
sequences, PARP homologues and antibodies are useful for analytic
detection of PARP homologues and for identifying PARP effectors or
binding partners, as well as for determining their effectiveness.
PARP-binding partners are useful for the diagnosis or therapy of a
disease condition, which is the result of a PARP protein, especially an
energy deficiency, which may comprise tissue damage from cell death
following necrosis or apoptosis, the disease condition may be chosen
from a neurodegenerative illness, or sepsis or ischemic tissue damage,
in particular neurotoxic disturbances, etc. This sequence encodes the
human PARP2 protein used in the method of the invention.
Sequence 1843 BP; 599 A; 377 C; 447 G; 420 T; 0 other;

alignment_scores:
Quality: 851.50 Length: 499
Ratio: 2.534 Gaps: 15
Percent Similarity: 67.335 Percent Identity: 40.882
alignment_block:
US-09-236-995D-2 x AAZ44287 ..
Align seg 1/1 to: AAZ44287 from: 1 to: 1843
519 GlnCys***LeuLysHisValLeuThr***His***ValCysThrGly.. 534
: : : : :
252 GAGTGTACAGCCAAG...GTGGGAAGGCTCATCTGTATTGTGAAGGAAA 298
: : : : :
534 534
299 TGATGCTCTATGATGTCATGCTAAATCAGACCAATCTCCAGTTCACAAACA 348
535TyrTyrValLeuGlnIleLeuGlnAspGlySerGlyCys 549
: : : : :
349 ACAAGTACTACTCTGATGCTATTAGAGATGATGATGATGATGATGATGAT 398
: : : : :
550 TyrValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlyG1 566
: : : : :
seq_documentation_block:
ID AAZ44287 standard; cDNA; 1843 BP.
XX
AC AAZ44287;
XX
DT 31-MAR-2000 (first entry)
XX
XX Human brain PARP2 cDNA.
XX
KW PARP; poly(ADP-ribose) polymerase; human; murine; detection; sepsis;
diagnosis; therapy; necrosis; apoptosis; neurodegenerative illness;
KW

399 AGTCTTTGGATGAGATGGGCGGAGTTGGG.....AAAATGGGACAGCA 442
566 nLysLeuGluMetSer.....LysThrGluAlaIleLysGluPheL 581
443 CAGCTGGTGGCTGTTTCAGGCAATCTCAACAGGCCAAGGAATCTTTC 492
581 ysArgLeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCys 597
493 AGAAGAAATTCCTTGCACAAACAAACAAACAAATTTGGGAAGATCGAGAA... 539
598 LysThrAsnPheArgLysGlnProGlyArgPheThrProLeuAspValas 614
540AAGTTGAGAGGTCCTGGAAATATGATATGATACAGATGGA 583
614 pTyrGlyValLysAlaProLysArgLysAspIleSerGlu.....M 629
584 CTATGCCCACTACTCAGGATGAAGAGGAACAAAGAAAGAGGAATCTC 633
629 etLysSerLeuAlaProGln.....LeuLeuGlu 639
634 TTAATCTCCCTTTGAAGCCAGAGTACAGCTAGATCTTCGGGTACAGGAG 683
640 LeuMetLysMetLeuPheAsnValGluThrTyrArgAlaAlaMetMetG1 656
584 TTAATAAGTTGATCTCTAATGTTAGGCCATGGAGAAATGATGATGGA 733
656 uPheGlu***AsnMetSerGluMetProLeuGlyLysLeuSer*****A 673
734 AATCAAGTATATACCAAGAACGCCACCTTGGGAAGCTGACAGTGGCAC 783
673 snIleGlu***GlyPheGluAlaLeuThr*****LeuPheGlu 689
784 AAATCAAGGAGGTTACCATCTCTTAAGAAGATGAGGATTTGATTCGG 833
690 GlyHisArg***SerSerThrGlyLeu***GluLysAla***LeuLeu** 706
834GCTGGCCAGCATGGAGAGCTCTCATGGAAGC 865
706 *****PheSerLeuLeuSerLeuLeuPheIleLeu..... 720
866 ATGCAATGAATTTACACAGGATTCGGCATGACTTTGGACTCCGTACTC 915
721IleLeuTyrGlyMetArgMetIleSerTyrSerLysAlaLysMet 735
916 CTCACCTAATCCGACACAGAGAACTGTCA...GAAAAATACAAATTA 962
736 LeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleValGlyPheAs 752
963 CTAGAGGCTTTGGGAGACATTTGAATTTGATTAAGCTGGTGAACACAGA 1012
752 pSerAspSerAspGlu...SerLeuAspAspLysTyrMetLysLeuHisc 768
1013 GCTACAAAGCCAGACACCCATTTGGACCAACACTATAGAACTACATTT 1062
768 ysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeuIleGlu 784
1063 GTGCTTGGCCCTTGACCATGAAAGTTACGAGTTCAAGATGATTC 1112
785 GlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTrpSerLeuG1 801
1113 CAGTACCTACAACTACCCATGCTCCACACACAGCGACTATACCATGAC 1162
801 uLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTyrS 818
1163 CTGTGCTGAGTTGTTTCAAGTGGAGAGATGTTGAG.....AAG 1203
818 erArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlySer 834
1204 AAGCCTTCAGAGAGGACCTTCATACAGAGGATGCTTCTATGGCATGTTCC 1253
835 ArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAlaPr 851
1254 AGGATGAGTAATCGGTGGGAATCTTTAGGCCATGGGCTTCGACATGCC 1303

851 oProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheA 868
1304 ACCTGAAGTCCCATCAGAGTTACATGTTGGAAAGGAATCTACTTTG 1353
868 laAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArgAsnAsn 884
1354 CTGACATGCTTCCCAAGAGTGCATTAAGTCTTGCCTTCGCTGCTAAAG 1403
885 ProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMetTyrG1 901
1404 ATACAGGACTGCTGCTCTTATCAGAGGTAGCTTAGGTCAAGTAAATGA 1453
901 uLeuLysLysAlaThr...SerMetAspLysProArgGlyLysHiss 917
1454 ACTAGAGGCCAATCTTAAGCCGAGGATGCTTCAAGGTAAACATA 1503
917 erThrLysGlyLeuGlyThrValProLeuGluSerGluPheValLys 933
1504 GCACCAAGGGGTGGCAGATGGCTCCAGTTCGCCCACTTCGTCACC 1553
934 TrpArgAspValValProCysGlyLysProValProSerSerI1 950
1554 CTGAATGGGAGT...ACAGTGCCATTAGCACCAGCAAGTCACAGGAAT 1600
950 eArgSerSerGlu.....LeuMetTyrAsnGluTyrIleValTyrA 964
1601 TCTGAATCCAGATGTTATACCTCACTACATCAATATATATATATA 1650
964 snThrSerGlnValLysMetGlnPheLeuLeuLysValArgPheHis 979
1651 ACCCAACAGGTCGATGCGGTACCTTTTAAAGGTTTCAGTTAAT 1697

seq_name: /SID52/gcdata/geneseq/geneseq/NA2001.DAT.AAC82090

seq_documentation_block:

ID AAC82090 standard; cDNA; 1843 BP.

XX AAC82090;

DT 02-MAR-2001 (first entry)

XX Human brain poly-ADP-ribose-polymerase cDNA.

XX Human; poly ADP-ribose polymerase; PARP; neuroprotective; neurotropic; cerebroprotective; antiparkinsonian; nephrotropic; cardiant; vasotropic; anticonvulsant; cytostatic; antibacterial; immunosuppressive; treatment; antiinflammatory; antirheumatic; antiarthritic; antidiabetic; epilepsy; 2H-phthalazin-1-one derivative; neurodegenerative disease; tumor; neuronal damage; Alzheimer's disease; Huntington's disease; metastasis; Parkinson's disease; ischemic damage; microinfarction; sepsis; diabetes mellitus; ss.

XX Homo sapiens.

XX DE19921567-A1.

XX 16-NOV-2000.

XX 11-MAY-1999; 99DE-1021567.

XX 11-MAY-1999; 99DE-1021567.

XX (BADI) BASF AG.

XX Lubisch W, Sadowski J, Kock M, Hoeger T;

XX WPI; 2001-032983/05.

XX P-PSDB; AAB11480.

XX Drugs for inhibiting PARP or especially homologous enzymes comprising 4-substituted phthalazinone derivatives, useful e.g. for treating neurodegenerative disease, ischemic damage, tumors or diabetes -

yy

Example A; Page 9-12; 14pp; German.

This invention describes novel 4-substituted 2H-phthalazin-1-one derivatives (I) which are used for the treatment or prophylaxis of diseases associated with elevated poly-(ADP-ribose)-polymerase (PARP; EC 2.4.2.30) activity. The products of the invention have neurotropic, neuroprotective, cerebroprotective, antiparkinsonian, nephrotropic, cardiatic, vasotropic, anticonvulsant, cytostatic, antibacterial, immunosuppressive, antiinflammatory, antirheumatic, antiarthritic, antididiabetic. (I) are especially used for treating or preventing neurodegenerative disease or neuronal damage (specifically associated with ischemia, trauma or massive bleeding, especially apoplexy or spinal-cranial trauma; or Alzheimer's disease, Huntington's disease or Parkinson's disease), treating or preventing ischemic damage (specifically renal damage after renal ischemia or during and after kidney transplantation or heart damage after cardiac ischemia), treating epilepsy, specifically generalized epileptic attacks (e.g. petit mal and tonic-clonic attacks) or partial epileptic attacks (e.g. temporal lobe and complex partial attacks), treating microinfarction (e.g. during and after heart valve replacement, aneurysm resectioning and heart transplantation), revascularization of critically constricted coronary arteries (e.g. after PCTA or by-pass operations) or peripheral arteries (e.g. leg arteries), treating acute myocardial ischemia and damage during or after its mechanical or drug-induced lysis and treating tumors and their metastasis, sepsis and septic shock, inflammatory and rheumatic disease (e.g. rheumatoid arthritis), and diabetes mellitus. Although (I) inhibit PARP (i.e. the known form designated PARP1), they especially selectively and strongly inhibit PARP homologs, specifically the homolog PARP2. In particular (I) have very strong PARP2 inhibitory activity (e.g. with K_i values of 1-20 nM) and high selectivity for PARP2 relative to PARP1 (generally by a factor of more than 5).

Sequence 1843 BP: 599 A: 377 C: 447 G: 420 T: 0 other;
XX

alignment scores:

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 851.50 | Length: 499 |
| Ratio: | 2.534 | Gaps: 15 |
| Percent Similarity: | 67.335 | Percent Identity: 40.882 |

alignment_block:

US-09-236-995D-2 x AAC82090

Align seg 1/1 to: AAC82090 from: 1 to: 1843

519 GlnCys***LeuIysHisValLeuThr***His***ValCysThrGly. 534

219 GAGGCTGGTGTATTTGTAAGCGAAA 298
:::||||| ||| ||||| ||| |||
GAGGCTGGTGTATTTGTAAGCGAAA 298

534 534

299 TGATGTCATGATGCTAAATCAGACCAATCTCCAGTTCACACAACA 348

535TyrTyrValLeuGlnIleIleGluGlnAspAspGlySerGluCys 549

349 ACAAGTACTATCTGATTGATTCAGCTATTAGAAGATGATGCCAGAGGAACCTC 398

550 TyrvalPheArgLysTrpGlyArgvalGlySerGluLysIleGlyGlyG1 566

399 AGTGTGGATGAGATGGGGCCGAGTTGGG.....AAAATGGGACAGCA 442

566 nLysLeuGluGluMetSer.....LysThrGluAlaIleLysGluPheL 581

443 CAGCCTGGTGGCTTGTTTCAGGCAATCTCAACAAGGCCAAGGAAATCTTTC 492

581 ysArgLeuPheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCys 597
 |||||..... |||||..... |||||..... |||||.....

493 AGAAGAAATTCCTTGACAAACGAAAAACAATGGGAAGATCGAGAA... 539

598 LysThrAsnPheArgLysGlnProGlyArgPheTyrProLeuAspValAs 614

540AAGTTTGAGAAGGTCCTGGAAAAATATGATATGCTACAGATGGA 583

614 pTyrGlyValLysLysAlaProLysArgLysAspIleSerGlu.....M 629
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
584 CTATGCCACCAATACTCAGGATGAAGAGGAAACAAGAAGAAAGGAATCTC 633

629 eTysSerSerLeuAlaProGln.....LeuLeuGlu 639
:||| ||| ||| ||| :||| :||| :||| :||| :||| :|||
634 TTAATCTCCCTTGAAGCCAGAGTCACAGCTAGATCTTCGGGTACAGGAG 683

640 LeuMetLysMetLeuPheAsnValIgluThrTyTargAlaAlaMetMetGl 656
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
684 TTAATAAAGTTGATCTGTAATGTTTCAGGCCATGGAACAATAATGATGGA 733

656 uPheGlu***AsnMetSerGluMetProLeuGlyLysLeuSer*****A 673
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
734 AATGAAGTATATAACCAAGAAAGCCCACATTGGGAAGCTCACAGTGCGAC 783

673 snIleGlu**GlyPheGluAlaLeuThr*****LeuPheGlu 689
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
784 AAATCAAGGCAGTTACCAGTCCTTAAAGAGATTGAGGATTGTATTTCGG 833

690 GlyHisArg***SerSerThrGlyLeu**GluLysAla***LeuLeu** 706
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
834GCTGGCCAGCATGGAGCACTCTCATGGAAGC 865

706 *****PheSerLeuLeuSerLeuLeuPheIleLeu..... 720
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
866 ATGCATGAATTCACACCAAGGATTCGCATGACTTTGGACTCCGTACTC 915

721IleLeuTyrglyMetArgMetIleSertyrSerLysAlaLysMet 735
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
916 CTCACACTAATCCGGACACACAGGAAGCACTGTCA...GAAAAAATACAAATTA 962

736 LeuGluAlaLeuGlnAspIleGluIleAlaSerLysIleValGlyPheAs 752
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
963 CTACAGGCTTTGGGAGACATTGAAATTGCTATTAAAGCTGGTGAACACAGA 1012

752 pSerAspSerAspGlu...SerLeuAspAspLysTyrrMetLysLeuHisc 768
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1013 GCTCAAAGCCCAACACCCATTGGACCAACACTATAGAACCTACATT 1062

768 ysAspIleThrProLeuAlaHisaspSerGluasptYrLysLeuIleGlu 784
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1063 GTGCCCTTCGCCCCCTTCACATCAAGATTCAGAGTTCAAAGATGATTTCC 1112

785 GlntYrLeuLeuAsnThrHisAlaproThrsLysAspTrpserLeuGl 801
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1113 CAGTACCTTCAATCTACCCATGCTCCACACACAGCAGCTATACCATGAC 1162

801 uIleGluGluValPheSerLeuAspAr-gAspgLygLuLeuAsnLysTyrS 818
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1163 CTTGCTGGATTGTTGAAGTGGAAGAGATGGTGAG.....AAG 1203

818 erArgTyTLySAsnAsnIleuHisasnLysMetLeuLeuTrpHisGlySer 834
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1204 AAGCCTTCAGAGAGGACCTTCATACAGAGATGCTTCTATGCGATGGTTCC 1253

835 ArgLeuThrAsnPhenValGlyIleLeuSerGlnGlyLeuArgileAlaPr 851
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1254 AGGATGAGTAACCTGGTGGGAATCTTGAGCCATGGGCTTCGAATTGCCCC 1303

851 oProGluAlaproValthrGlyTyrrMetPheGlyLysGlyLeuTyrrPhea 868
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1304 ACCTGAAGCTCCCATCACAGGTTACATGTTTGGGAAGGAATCTACATTG 1353

868 IaAspLeuValSerLysSerAlaGlnTyrcYstYrValaspArgAsnAs 884
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1354 CTGACATGCTCTCCNAGAGAGTCCCAATTACTGCTTTGCCCTCGCCTAAAG 1403

885 ProvalGlyLeuMetLeuLeuSerGluValAlaLeuGlyLysAspMetTyrgl 901
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1404 AATACAGACTGCTGCTCTTATACAGAGTAGCTCTAGTCACTAGTGAATGA 1453

950 eArgSerSerGlu.....LeuMetTyrAsnGluTyrIleValTyrA 964
 165 TCTGAATCCAGATGGTTATACCCCTCAACTCAATGAATATATGTTATAT 1714
 964 snThrSerGlnValLysMetGlnPheLeuLeuTysValArgPheHis 979
 1715 ACCCAACACCGATCCGATATGCGGTACCTTTTAAAGGTTTCAGTTTAAAT 1761

309 AGTCGCCCATGTTTGGTGAAGGACCTCCACACTGGTACCACCTTCTCTGCG 358
56 IlePheSerLysLysAsnGlnIleLysSerValAsp...AspValGluGI 71
359 TTCTGGAGGTGGCCCATCTCCATCCGGACCCCTGACGTTGAGGTGGATGG 408
71 yileAspAlaLeuArgTrpAspAspGlnGluLysIleArgAsnTyrVal. 87
409 GTTCTCTGAGCTTGGTGGGTGACACAGAGTCAAGAGTCAAGAGACAGCGG 458
88GlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
459 AAGCTGGAGGATGACAGCAAGCCAGGATGGAAATGGTAGCAAGCA 508
102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
509 GAGAGACTCTGGGTGACTTTCACAGCAGAGTATGCCAAGTCCACAGAG 558
114 rSerCysArgArgCysSerGluLysIleThrLysGlySerValArgLeuS 131
559 TAGCTCAAGGGGTGTATGGAGAAGATAGAAAAGGGCCAGGTGCCCTGT 608
131 eAlaLysLeu...GluSerGluGlyProLys...GlyIle.....Pro 143
609 CCAAGAAGATGGTGGACCGGAGAGCCACACTAGGCATGATTGACCGC 658
144 TrpTyrHisAlaAsnCysPhePheGlu.....ValSe 154
659 TGCTACCATCCAGGCTGCTTTGTCAAGAACAGGAGGAGCTGGGTTCGG 708
154 rProSerAlaThrValGluLysPheSerGlyTrpArgThrLysSerAspG 171
709 GCCCGAGTACAGTGCAGGACGAGTCAAGGGCTTCAGCCTCTTGTCTAC 758
171 luAspLysArgThrMetLeuAspLeuValLysLysAspValGlyAsnAsn 187
759 AGGATAA.....GAAGCCCTGAAGAAGCAGCTCCAGGAGTC 796
188 GluGlnAsnLysGlySerLysArgLysLysSerGluAsnAspIleAspSe 204
797 AAG.....AGTGAAGAAAGAGAAAGAGGCGATGAGTGGATGG 834
204 rTyrLysSerAlaLeuAspLysSerThrSerGluGlyThrValArgA 221
835 A.....GTGGATGAGTGGCGAAGAGAAATCTAATAAAG 869
221 snLysGlyGlnLeuValAspProArgGlySerAsnThrSerSerAlaAsp 237
870 AAAAGACAAG.....GATAGTAAG 889
238 IleGlnLeuLysLysLysGluGlnSerAspThrLysLysLysLysAs 254
890 CTTGAAAAGCCCTAAAGGCTCAGACGACCTGATCTGGAACATCAAGGA 939
254 pGlyLeuLysThrHisValSerAlaGluLeuArgAspMetLeuGluA 271
940 CGAGCTAAAGAAAGTGTGTCACTAATGACCTGAAGGAGCTACTCATCT 989
271 laAsnGlyGlnAspThrSerGlyProGluArgHisLeuLeuAspArgCys 287
990 TCAACAAGCAGCAAGTCCCTTGGGAGGTCGGGATCTTGGACCGAGTA 1039
288 AlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAlaAs 304
1040 CTTGATGGCATGGTTCGGTGGCCCTCTTCCCTGCGAGGAATGCTCGGG 1089
304 nGlyMetTyrTyrAsnGlyGlnThrGlnCysSerGlyAsnValSerG 321
1090 TCAGCTGGTCTTCAAGACGAGTGGCTTATCTGCACTGGGAGCTCACTG 1139
321 luTrpSerLysCysThrTyrSerAlaThrGluProValArgValLysLys 337
1140 CTGGACCAAGTGTATGCTCAAGACACAGACACCAACCCG.....AAG 1183

338 LysTrpGlnIleProHisGlyThrLysAsnAspTyrLeuMetLysTrpPh 354
1184 GAGTGGGTAAACCCA.....AAGGAATCCGAGAAATCTCTTACCT 1224
354 eLysSerGlnLysValLysLysProGluArgValLeuProProMetSer. 370
1225 CAAGAAATTCAGGTAAAAAGCAGGACCGTATATTCCTCCCAAGAACCA 1274
371ProGluLysSerGlySerLysAla 378
1275 GCGCTCCGTGGGCGCACGCTCCCGCTCCACAGCCTCGGCTCTGCT 1324
379 ThrGluArgThrSerLeuSerSerLysGlyLeuAspLysLysLeuArgPh 395
1325 GCTGTGAACCTCTCTGCTCAGCAGATAAGCCATTATCCAAACATGAAGAT 1374
395 eSerValValGly.....GlnSerLysGluAlaAlaAsnGluTrpI 409
1375 CCTGACTCTCGGAAGCTGTCCCGGAACAAGATGAAGTGAAGGCCCATGA 1424
409 leGluLysLeu.....LysLeuAlaGlyAlaAsnPheTyrAlaArgVal 423
1425 TTGAGAACTCGGGGGAAGTGTACGGGAGCGGCCAACACAGGCTTCCCTG 1474
424 VallysAspIleAspCysLeuIleAlaCysGlyGluLeuAspAsnGluAs 440
1475TGCATCAGCACCAAAAGGAGGTGGAAGAGATGAA 1509
440 nAlaGluValArgLysAlaArgLeuLysIleProIleValArgGluG 457
1510 TAAGAAGATGGAGGAAGTAAAGGAAGCCACATCCGAGTTGTGTCTGAGG 1559
457 lyTyrIleGlyCysValLysArgThrLys..... 467
1560 ACTTCTCCAGGAGCTCTCCGCTCCACCAAGAGCCTTCAGGAGTTGTC 1609
468 CysCysHisLeuIleCysIleAsnTrpAsnAlaLeuGluSerLysG1 484
1610 TTAGCGCATCTTGTGCC...CCTTGGGGGCGCAGAGGTGAAGCGCAGGCC 1656
484 y**ThrValThrValLysValLysGlyArgSerAlaCysSer***SerP 501
1657 TGTGAAGTT...GTGGCCCCAAGAGGAACTCAGGGCTCGCTCTCCA 1703
501 ro**ValCysLysAsnThrAlaHisIlePro*****Trp..... 514
1704 AA.....AAAAGCAAGGCCAGGTCAAGAGGAGGTATCAACAAA 1744
515 ...GluLysHisIleGlnCys***LeuLys..... 523
1745 TCTGAAAGAGAAATGAAATTAATCTTTAAAGAGGAGCAGCAGCTGTGATCC 1794
524HisValLeu..... 526
1795 TGATTTGGACTGGAACACTCTCGCGCATGTCTCGGAGAAAGGTGGGAAGG 1844
527Thr***His***Val.....CysThrGly 534
1845 TCTTCAAGTCCACCTTGGCTGGTGGACATCGTTAAAGGAACCACTCC 1894
535 TyrTyrValLeuGlnIleGluGlnAspAspGlySerGluCysTyrVa 551
1895 TACTACAAGCTCGAGCTTCTGAGGAGCAGCAAGGAACACAGGTATGGAT 1944
551 lPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLysL 568
1945 ATTCAAGTCTCGGGCGGTGTGGGTACG...GTGATCGGTAGCAACAAAC 1991
568 euGluGluMet...SerLysThrGluAlaIleLysGluPheLysArgLeu 583
1992 TGGACAGCATGCGGTCCCAAGGAGGATGCCATTGACAGGTTTCATGAATTA 2041

584 PheLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAs 600
2042 TATGAAGAAAACCGGAGAGCTTGGCAGTCC.....AAAAA 2079
600 nPheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrGlyV 617
2080 TTTACGAGATATCCCAAAAGATTTTACCCCTGGAGATTGACTATGGCC 2129
617 al...LysLysAlaProLysArgLysAspLysSerGlu...MetLysSer 631
2130 AGGATGAAGAGGAGCTGAAGAAGCTCAGATAAATCTCGGCCCAAGTCC 2179
632 SerLeuAlaProGlnLeuLeuGluLeuMetLysMetLeuPheAsnValGI 648
2180 AAGCTCCCAAGCCAGTTCAGGACCTCATCAAGATGATCTTGTATGGGA 2229
648 uThrTyrArgAlaAlaMetMetGluPheGlu***AsnMetSerGluMetp 665
2230 AAGTATGAAGAAACCCATGGTGGAGTATGAGATCGACCTTCAGAAAGATGC 2279
665 roLeuGlyLysLeuSer*****AsnIleGlu**GlyPheGluAlaLeu 681
2280 CCTTGGGAGAGCTGAGCAAAAGGAGATCCAGGCGGCATATCCATCCCTC 2329
682 Thr*****LeuPheGluGlyHisArg***SerSerThrGlyLe 698
2330 AGT.....GAGGTCCAGCAGCGGTGTCTCAGGCGAG 2361
698 u**GluLysAla***LeuLeu*****PheSerLeuLeu 715
2362 CAGGACTCTCAGATCTCGATCTCAATCGCTTTTACACCTGATCC 2411
715 erLeuLeuPheIleLeuLeuTyrGlyMetArgMet..... 727
2412 CCCACGACTTT.....GGGATGAAGAAGCTCCGCTCCCTG 2446
728IleSerTyrSerLysAlaLysMetLeuCluAlaLeuGl 740
2447 AACATGACACAGTGTGAGGCGAGGTGAAATGCTTGACAACTGCT 2496
740 nAspIleGluIleAlaSerLysIleVal.....GlyPheAspSerAsps 755
2497 GGACATCGAGGTGCTTACAGTCTGCTCAGGCGAGGTCTGATGATGACA 2546
755 erAspGluSerLeuAspAspLysTyrMetLysLeuHisCysAspIleThr 771
2547 GCAAGATCCCATCGATGTCAACTATGAGAAGCTCAAACTGACATTAAG 2596
772 ProLeuAlaHisAspSerGluAspTyrLysLeuIleGluClnTyrLeuLe 788
2597 GTGTTGACAGAGATTCTGAAGAAGCCGAGATCATCAGGAAGTATGTTAA 2646
788 uAsnThrHisAlaProThrHisLysAspTrpSerLeuGluLeuGluV 805
2647 GAACACTCATGCAACACACACATCGGTATGACTTGGAGTCAATCGATA 2696
805 alPheSerLeuAspArgAspGlyGluLeuAsnLysTyrSerArgTyrLys 821
2697 TCTTTAAGATAGCGCTGAAGCGAATGCGAGCGTTACAGCCCTTTAAG 2746
822 AsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAs 838
2747 ...CAGCTTCATAACCGAAGATGCTGTGGCAGCGGTCCAGGACCAACAA 2793
838 nPheValGlyIleLeuSerGlnGlyLeuArgIleAlaProGluAlaAp 855
2794 CTTTGTGGATCTCTCCAGGGTCTTGGGATAGCCCGCGCTGAAGCGC 2843
855 roValThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuVal 871
2844 CCGTGACAGGCTACATGTTTGTGAAGGATCTATTTCGCTGACATGCTC 2893
872 SerLysSerAlaGlnTyrCysTyrValArgAsnAsnProValGlyLe 888

2894 TCCAAGAGTGCCAACTACTACCATAGTCTCAGGAGACCCCAATAGGCTT 2943
888 uMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysA 905
2944 AATCCCTGTTGGGAAAGTTCCCTTGGAAACATGTATGAAGTGAAGCAG 2993
905 laThrSerMetAspLysProProArgGlyLysHisSerThrLysGlyLeu 921
2994 CTTACATATCAGCAGGTTACCCAGGAGGACACAGTGTCAAGGTTTG 3043
922 GlyLysThrValProLeuGluSerGluPheValLysTrpArgAspVa 938
3044 GGCAGAACTACCCCTGATCTCTCAGCTAACATAGT...CTGATGGTCT 3090
938 lValValProCysGlyLysProValProSerSerIleArgSerSerGluL 955
3091 AGAGTTCCTCTTGGGACCGGATTCATCTGCTGTAATGACACCTCTC 3140
955 euMetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGln 971
3141 TACTATATAACAGTACATGTCTATGATATTGCTCAGGTAATCTGAAG 3190
972 PheLeuLeuLysValArgPheHisLys 981
3191 TATCTGCTGAAACTGAAATCAATTTAAG 3220
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-7

seq_documentation_block:

; Sequence 7, Application US/08044618
; Patent No. 5449605

; GENERAL INFORMATION:

; APPLICANT: SMULSON, MARK

; TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO

; TITLE OF INVENTION: CANCER BY THE USED OF RESTRICTION FRAGMENT LENGTH

; TITLE OF INVENTION: POLYMORPHISM OF THE GENE FOR THE HUMAN POLY (ADP-RIBOSE)

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1225 Connecticut Suite 300

; CITY: Washington

; STATE: D.C.

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/044,618

; FILING DATE: 19930406

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/257,696

; FILING DATE: 14-OCT-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: FOX, SAMUEL L

; REGISTRATION NUMBER: 30,353

; REFERENCE/DOCKET NUMBER: 0654.0490001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)466-0800

; TELEFAX: (202)833-8716

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5345 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: both

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

; US-08-044-618-7

```

alignment_scores:      Quality: 1237.00      Length: 1128
                       Ratio: 1.811      Gaps: 49
                       Percent Identity: 33.511
Percent Similarity: 60.550
alignment_block:
US-09-236-995D-2 x US-08-044-618-7  ..
Align seg 1/1 to: US-08-044-618-7  from: 1 to: 5345
6  LysAlaThrLysAlaGluTyralAlaLysSerGlyArgAlaSerCysLysSe 22
   |||   :::   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
791 AAGCTCTACTGAGTGAGCTACACCAAGAGCGGCGCACCTCTTGTGAAGAA 840
22 rCysArgSerProIleAlaLysAspGlnLeuArgLeuGlyLysMetValG 39
   |||   :::   |||   |||||:|||||:|||||:|||||:|||||:|||||
841 ATGCAGCAGGACATCAACCCAGGATCGCTCCGATGGCCATCATGTGTC 890
39 InAlaSerGlnPheAspGlyPheMetProMetTrpAsnHisAlaArgCys 55
   ||   |||:|:|   |||   |||   |||   |||   |||   |||   |||
891 AATATGTCATGTT.GATGAGAAGTCCCAACACTGGTACCACCTTCTCCTG 939
56 IlePheSerLysLysAsnGlnIleLysSerValAsp...AspValGluGl 71
   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
940 TTGCAGAAGGTGGCGCACCTCCATGGCGGCACCTGACGTGTGAGTGGATGG 989
71 yIleAspAlaLeuArgTrpAspAspGlnGlyLysIle.....ArgA 85
   ||:|:|:|:|   |||   |||   |||   |||   |||   |||   |||
990 ATTCTCTGAGCTTTGGTGGGATGACCAGCAGAAAGTCAAGAAGACACGCTG 1039
85 snTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
   |||   |||   ::::|   |||   |||   |||   |||   |||   |||
1040 ACGCTGGAGAGTGACGGGCAAGGCCAAGATGGAATGGTGTAGCAAGGCA 1089
102 GluLysCys.....ThrIleGluIleAlaProSerAlaArgTh 114
   |||||   |||   |||   |||   |||   |||   |||   |||   |||
1090 GAGAAGACCTGGGTGACTTTCAGCAGAGATAGGCCAAGTCCAGCAGAAG 1139
114 rSerCysArgArgCys.SerGluLysIleThrLysGlySerValArgLeu 130
   |||   |||:|:|   |||   |||   |||   |||   |||   |||   |||
1140 CAAGTGCAGAGAGTGTTAT7GGAGAAGATAGAAAGGGCCAGATGCGCCTA 1189
131 SerAlaLys...LeuGluSerGluGlyProLys...GlyTle.....Pr 143
   |||   |||   |||:|   |||   |||   |||   |||   |||   |||
1190 TCCAAGAAGATGCTGGACCTGAGAAGCCTCAGCTAGGCATGATTGACCC 1239
143 oTrpTyrHisAlaAsnCysPhePheGlu.....ValS 154
   |||   |||:|:|   |||   |||   |||   |||   |||   |||   |||
1240 CTGGTACCACCCAGGCTGCTTTTCTCAAGAACAGGGAGGAGCTGGGTTC 1289
154 exProSerAlaThrValGluLysPheSerGlyTrpAspThrLeuSerAsp 170
   :|||:|:|   :|||:|:|   :|||:|:|   :|||:|:|   :|||:|:|
1290 AGCTGAGTACAGTGGGATCAGCTCAGGGCTTCAGCTCTCGTGCA 1339
171 GluAspLysArgThrMetLeuAspLeuValLysLysAspValGlyAsnAs 187
   |||||:|:|   |||:|:|   |||:|:|   |||:|:|   |||:|:|
1340 GAGGATAAAGAAACCCCTG.....AAGAAGCAGCTCCAGGAGT 1377
187 nGluGlnAsnLysGlySerLysArgLysLysSerGluAsnAspIleAspS 204
   :|||:|:|   |||:|:|   |||:|:|   |||:|:|   |||:|:|
1378 CAAG.....AGTGAAGAAAGAGAAGGCGATGAGGTGGATG 1415
204 erTyrLysSerAlaArgLeuAspGluSerThrSerGluGlyThrValArg 220
   ::   :|||:|:|   |||:|:|   |||:|:|   |||:|:|   |||:|:|
1416 GA.....GTGGATGAA.....GTGGCCAAG 1435
221 AsnLysGlyGlnLeuValAspProArgGly.SerAsnThrSerSerAlaA 237
   :|||:|:|   |||:|:|   |||:|:|   |||:|:|   |||:|:|
1436 AAGAAATCTAAAAAATAAAAAAACGAGGATATAAGCTTGAAAGGCC. 1484
237 spIleGlnLeuLysLeuLysGluGlnSerAspThrLeuTrpLysLeuLys 253
   |||||   |||   |||   |||   |||   |||   |||   |||   |||

```

[illegible]

506 snThrAlaHisIlePro*.....Trp..... 514
||||| : : : : : |||||
2385 ACACCTCTGCGCATG.....TCCTGGAGAAAGGTGAAGGCAGAGCCTGTT 2428
514 514
2429 GAAGTCGTAGCCCCAAGAGGAACTCAGGAGCTGTGCTCTCCCAAAAAAG 2478
515 GluLysHisIleG 519
2479 CAAGGGCCAGGTCAAGGAGGAGGATATCAAAATCTGAAGAACAGATGA 2528
519 InCys***LeuLys..... 523
2529 AATTAACTCTTAAAGGAGGAGGAGCAGCTGTGGATCTGACTCTGCTCTGGAA 2578
524HisValLeu.....Thr** 528
2579 CACTCTGCACATGTCGAGAGAGGTGGGAAGGTCTTCAGTGCCACCCCT 2628
528 *His***Val.....CysThrGlyTyrTyrValLeuGlnI 540
||| : : : : : ||||| : : : : :
2629 CAGCCTGGTGGACGTCGTTAAAGGAACCACTCTCTATTACAGCTGAAGT 2678
540 lelleGluInAspAspGlySerGluCysTyrValPheArgLysTrpGly 556
: : : : : ||| : : : : : ||||| : : : : :
2679 TGCTGAAGGATGACAGGAAGGAGGAGCAGCTGTGGATCTGACTCTGCTGGAC 2728
557 ArgValGlySerGluLysIleGlyGlnLysLeuGluMet...Se 572
||||| : : : : : ||||| : : : : : ||||| : : : : :
2729 CGTGTGGGCAG...GTGATGCGTATGACAACTGGAACAGCTGTGTC 2775
572 rLysThrGluAlaIleLysGluPheLysArgLeuPheLeuGluLysThrG 589
||||| : : : : : ||||| : : : : : ||||| : : : : :
2776 CAAGGAGGACACCATTCACACTTCATGAAATATATGAGAAAACTAG 2825
589 lyAsnSerTrpGluAlaTrpGluCysLysThrAsnPheArgLysGlnPro 605
|| : : : : : ||| : : : : : ||||| : : : : :
2826 GAAT.....GCTGGCAGCTCCAAAAA...TTCACAAAGTATCCC 2861
606 GlyArgPheTyrProLeuAspValAspTyrGlyVal...LysLysAlaPr 621
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
2862 AAAAAAGTTACCCCTCGAGATGAGTACGCGCCAGGACAAAGAGCGGT 2911
621 oLysArgLysAspIleSerGluMetLysSerSerLeuAlaProGln...L 637
||| : : : : : ||||| : : : : : ||||| : : : : :
2912 GAAGAGCCGACAGTAATCTCTGGCACCACCATGCTCCCAAGCCAG 2961
637 euLeuGluLeuMetLysMetLeuPheAsnValGluThrTyrArgAlaAla 653
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
2962 TTCAGGACCTCATCAAGATGATCTTTGATGTGGAAGTATGAGCAAGCC 3011
654 MetMetGluPheGlu***AsnMetSerGluMetProLeuGlyLysLeuSe 670
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
3012 ATGCTGGGTGTGAGATCAACCTT...CAGATGCCCTTGGGGAAGCTGAG 3058
670 r*****AsnIleGlu***GlyPheGluAlaLeuThr*****L 687
| : : : : : ||||| : : : : : ||||| : : : : :
3059 CAAAGGCAATCCAGGCGCGTACTCTCCATCTCAGGTC..... 3097
687 euPheGluGlyHisArg***SerSerThrGlyLeu***GluLysAla*** 703
: : : : : ||| : : : : : ||||| : : : : :
3098CAGCAGGTGGTGTCCAGGGGAGGAGGAGCTCTCAGATC 3136
704 LeuLeu*****PheSerLeuLeuSerLeuLeuPheIleLe 720
||| : : : : : ||||| : : : : : ||||| : : : : :
3137 Cys.....GATCTCTCAAAATCGCTTTTATCTCT 3165
720 uLeuLeu.....TyrGlyMetArg.....MetIleSer..... 729
||||| : : : : : ||||| : : : : :
3166 GATCCCCCAGCAGCTTTGGGATGAGGATCCTCTGCTCTGCTGGAACATGCGAG 3215

730TyrSerLysAlaLysMetLeuGluAlaLeuGlnAspIleGlu 743
: : : : : ||||| : : : : : ||||| : : : : :
3216 ACAGTGTGCGAGCCCAAGGTAGAATCTGGCAACACTGCTGGACATTTGAG 3265
744 IleAlaSerLysIleVal.....GlyPheAspSerAspSerAspGluSe 758
: : : : : ||||| : : : : : ||||| : : : : :
3266 GTAGCTACGGTCTGCTCAGGGGAGGCTTCCAGATAGCAGGAAGGACTC 3315
758 rLeuAspAspLysTyrMetLysLeuHisCysAspIleThrProLeuAlaH 775
| : : : : : ||||| : : : : : ||||| : : : : :
3316 CATCGATGTCACTATGACAAGCTCAAACTGACATTAAGGTGGTTGACA 3365
775 isAspSerGluAspTyrLysLeuIleGluGlnTyrLeuLeuAsnThrHis 791
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
3366 GAGATTTCTGAGGAAGCTGAGATCATCAGGAAGTATGTTAAGAACACTCAT 3415
792 Ala.Pro...ThrHisLysAspTrpSerLeuGluLeuGluValPheS 807
||| ||| ||||| : : : : : ||||| : : : : :
3416 GCAACCCACCCACACGATGATGACTTGGAGTCAATGATAGCTTTTA 3465
807 erLeuAspArgAspGlyGluLeuAsnLysTyrSerArgTyrLysAsnAsn 823
: : : : : ||| : : : : : ||||| : : : : :
3466 AGATAGAGTGTGAAGAGGAGTCCAGACTACAGCCCTTTAAG...CAG 3512
824 LeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheVa 840
||||| : : : : : ||||| : : : : : ||||| : : : : :
3513 CTTCACTAACTGAAGTCTGCTGGCATGGTCCAGGACCACCACTTGC 3562
840 lGlyIleLeuSerGlnGlyLeuArgIleAlaProProGluAlaProValT 857
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
3563 TGGGATCTGTCCTCCCTGGGTCTTTGGATAGCCCTGCTGGAACACCTGTGA 3612
857 hrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLys 873
||||| : : : : : ||||| : : : : : ||||| : : : : :
3613 TGGGTACATGTTGGTAAAGTATCTATTTCGCTGATCTTGTCTCCAAAG 3662
874 SerAlaGlnTyrCysTyrValAspArgAsnProValGlyLeuMetLe 890
: : : : : ||| : : : : : ||||| : : : : : ||||| : : : : :
3663 AGTGCAACGAGTGCATACATCTTAGGAAGACCAATAGGTTAATCCT 3712
890 uLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrS 907
| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
3713 GTCGGAAGAAGTTGCCCTTGGAAAGCTGTGTAACCTGAAGCATCTTCAC 3762
907 erMetAspLysProArgGlyLysHisSerThrLysGlyLeuGlyLys 923
: : : : : ||| : : : : : ||||| : : : : : ||||| : : : : :
3763 ATATCAGCAAGTTACCAAGGCGACACACAGTGTCAAGGTTTGGGCAAA 3812
924 ThrValProLeuGluSerGluPheValLysTrpArgAspValValVa 940
||||| : : : : : ||| : : : : : ||||| : : : : :
3813 ACTACTCTGACCTTTCAGCTAGTATCCCA...CTGGATGGTGTAGAGT 3859
940 lProCysGlyLysProValProSerSerIleArgSerSerGluLeuMetT 957
||||| : : : : : ||||| : : : : : ||||| : : : : :
3860 TCCTCTTGGGACCGAGGTTTCATCTGGTGTGAATGACACCTGTCTACTGT 3909
957 yrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLeu 973
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
3910 ATAATGATACATGCTCTATGATATTGCTCAGGTAATCTCAAAATATCTG 3959
974 LeuLysValArgPheHisLysLys 981
||||| : : : : : ||||| : : : : : ||||| : : : : :
3960 CTGAACCTGAAATTCATTTAAG 3983

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-3

seq_documentation_block:

; Sequence 3, Application US/08044618

; Patent No. 5449605

; GENERAL INFORMATION:

; APPLICANT: SMULSON, MARK

; TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO

; TITLE OF INVENTION: CANCER BY THE USED OF RESTRICTION FRAGMENT LENGTH

1170GAGCTGTTAAAGGAACCAACTCTCTATTACAGTGAAGT 1208
540 IleIleGluGlnAspAspGlySerGluCysTyrValPheArgLysTrpG1 556
1209 TTGCTGAAGGATGACAGGAAGACGAGCAITGGATATTCAAGTCTGGGA 1258
556 yArgValGlySerGluLysIleGlyGlnLysLeuGluGluMet...S 572
1259 CCGTGTGGGACG...GTGATCGGTAGCAACAACACTGGAACAGATGCTGT 1305
572 erLysThrGluAlaIleLysLysPheLysArgLeuPheLeuGluLysThr 588
1306 CCAAGGAGGACCACTGAACACTTATGAAATATATGAAGAAACTA 1355
589 GlyAsnSerTrpGluAlaTrpGluCysLysThrAsnPheArgLysGlnPr 605
1356 GGAAT.....GCTGGCACTCAAAAA...TTCACAAAGTATCC 1391
605 oGlyArgPheTrpProLeuAspValAspTyrGlyVal...LysLysAlaP 621
1392 CAAAAGTTCTACCCCTGGAGATTGACTACGGCCAGGACAAAGAGCGG 1441
621 roLysArgLysAspIleSerGluMetLysSerSerLeuAlaProGln... 636
1442 TGAAGAGCCGACAGTAACTCTGGCCACCAAGTCCATGCTCCCAAGCCA 1491
637 LeuLeuGluMetLysMetLeuPheAsnValIleThrTrpArgAlaAl 653
1492 GTTCAGGACCTATCAACAATGATCTTGATGTGAAAGTATGCAAGGC 1541
653 aMetMetGluPheGlu***AsnMetSerGluMetProLeuGluLysLeu 670
1542 CCATGGGGGTGTGAGATCAACCTT...CAGATGCCCTTGGGGAAGCTGA 1588
670 er****AsnIleGlu***GlyPheGluAlaLeuThr***** 686
1589 CAAAAGGCAATCCAGGCGCGTACTCCATCTCAGGTC..... 1628
687 LeuPheGluGlyHisArg***SerSerThrGlyLeu***GluLysAla** 703
1629CAGCAGGTGGTCTCCAGGCGCAGCAGCAGCTCAGAT 1666
703 *LeuLeu*****PheSerLeuLeuSerLeuPheIleL 720
1667 CCGT.....GATCTCTCAAAATCGCTTTTATATCC 1695
720 euIleLeu.....TyrGlyMetArg.....MetIleSer..... 729
1696 TGATCCCCCAGCAGTCTGGGATGAAGGATCCTCTGCTCTGGAACAATGCA 1745
730TyrSerLysAlaLysMetLeuGluAlaLeuGlnAspIleG1 743
1746 GACAGTGTGAGGCCAAGGTAGAAATGCTGGAACACTGCTGGCAGATTGA 1795
743 uIleAlaSerLysIleVal.....GlyPheAspSerAspSerGluS 758
1796 GGTAGCCTAGCGTCTGCTAGGAGGGGTCTCACGATAGCAGGAAGGACT 1845
758 erLeuAspAspLysTyrMetLysLysHisCysAspIleThrProLeuAla 774
1846 CCATCGATGCTCAATGAGAGCTCAAAACTGACATTAAGGTGGTTGAC 1895
775 HisAspSerGluAspTyrLysLeuIleGluLysThrLeuLeuAsnThrH1 791
1896 AGAGATTCTCAAGAAGCTGAGATCATCAGGAAGTATGTTAAGAACACTCA 1945
791 ala.Pro...ThrHisLysAspTrpSerLeuGluLeuGluValPhe 806
1946 TGCAACCAACACACACAGGATGATGATGCTTGAAGTCAATTGATGCTTT 1995
807 SerLeuAspArgAspGlyGluLeuAsnLysTyrSerArgTyrLysAsnAs 823
1996 AAGATAGAGTGTGAAGAGGAGTCCAGCAGCTACAAAGCCCTTTAAG...CA 2042

823 nLeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheV 840
2043 GCTCTATAACTGAAGGTGCTGTGGCATGGTCCAGGACCACCAACTTGT 2092
840 aIcylIleLeuSerGlnGlyLeuArgIleAlaProGluAlaProVal 856
2093 CTGGGATCTGCTCCCTGGGTCTTTGATAGCCCTGCTGCTGAAGCAGCTGTG 2142
857 ThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerly 873
2143 ATGGGCTACATGTTTGTAAAGTATCTATTTCGCTGATCTGTCTCCAA 2192
873 sSerAlaGlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetL 890
2193 GAGTGCAACGAGCTGCCATACATCTTAGGAAGACCAATAGGTTAATCC 2242
890 euLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThr 906
2243 TGTGCGAAGAAGTTGCCCTTGGAAACGTGTGTAACCTGAAGCATCTTCA 2292
907 SerMetAspLysProProArgGlyLysHisSerThrLysGlyLeuGly 923
2293 CATATCAGCAAGTTACCAAGGCGACGACAGTGTCAAGGTTTGGCAA 2342
923 sThrValProLeuGluSerGluPheValLysTrpArgAspValValV 940
2343 AACTACTCTGACCTTTCAGCTAGTATCCCA...CTGGATGGTGTAGAGG 2389
940 aIProCysGlyLysProValProSerSerIleArgSerSerGluLeuMet 956
2390 TTCTCTTGGGACCGAGGTTTTCATCTGTGTGTAATGACACTGTCTACTGT 2439
957 TyrAsnGluTyrIleValTyrAsnThrSerGlnValLysMetGlnPheLe 973
2440 TATATGAGTACATCTCTATGATATGCTCAGGTAAATCTGAAATATCT 2489
973 uLeuLysValArgPheHisHisLys 981
2490 GCTGAAACTGAAATTCATTTTAAG 2514

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-044-618-4

seq_documentation_block:

; Sequence 4, Application US/08044618
; Patent No. 5449605

; GENERAL INFORMATION:

; APPLICANT: SMULSON, MARK

; TITLE OF INVENTION: METHOD OR DETECTING A PREDISPOSITION TO

; TITLE OF INVENTION: CANCER BY THE USED OF RESTRICTION FRAGMENT LENGTH

; TITLE OF INVENTION: POLYMORPHISM OF THE GENE FOR THE HUMAN POLY (ADP-RIBOSE)

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1225 Connecticut Suite 300

; CITY: Washington

; STATE: D.C.

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/044,618

; FILING DATE: 19930406

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/257,696

; FILING DATE: 14-OCT-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: FOX, SAMUEL L

917GCGAGGTAGAAATGCTGGACA 938
738 laLeuGlnAspIleGluIleAlaSerLysIleVal.....GlyPheAsp 752
939 ACCTGCTGGACATGAGTAGCTACGCTGCTCGCAGGGAGGCTCTCAC 988
753 SerAspSerAspGluSerLeuAspAspLysTyrMetLysLeuHisCysAs 769
989 GATAGCAGGAGGAGCTCATCTGATGTCAACTATGAGAAGCTCAAACTGA 1038
769 pIleThrProLeuAlaHisAspSerGluAspTyrLysLeuIleGluCln 786
1039 CATTAAGCTGTTGACAGAGATCTGAAGAAGCTGAGATCATCAGGAAGT 1088
786 yrLeuLeuAsnThrHisAla.Pro...ThrHisLysAspTrpSerLeuG1 801
1089 ATGTTAAGAACACTCATGCAACCAACACACACGATGCATATGACTTGA 1138
801 uLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTyrS 818
1139 AGTCATTGATAGCTTTAAGATAGAGTGTGAAGAGAGAGTGCACACTACA 1188
818 erArgTyrLysAsnAspLeuHisAsnLysMetLeuLeuTrpHisGlySer 834
1189 AGCCCTTAAG...CAGCTTCAATACTGAAGGTGCTGTGGCATGGGTCC 1235
835 ArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAlaPr 851
1236 AGGACCACCAACTTGTGGGATGCTTCTGCTGGTCTTTGGATAGCCCT 1285
851 oProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheA 868
1286 GCCTGAAGCAGCTGCTGATGCTGATGCTGATGCTTGTGTAAGTATCTTCG 1335
868 laAspLeuValSerLysSerAlaClnTyrCysTyrValAspArgAsnAsn 884
1336 CTGATCTGTCTCAAGAGTGCACAGCAGCTGCATACATCTTAGGAAGAC 1385
885 ProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMetTyrG1 901
1386 CCAATAGGTTAATCTCTGCGAAGAGTTCCTTGGAAACGCTGTGA 1435
901 uLeuLysLysAlaThrSerMetAspLysProProArgGlyLysHisSerT 918
1436 ACTGAGCATGCTTCATATACAGCAAGTTACCCAAAGGCAAGCAGCATG 1485
918 hrLysGlyLeuGlyLysThrValProLeuGluSerGluPheValLysTrp 934
1486 TCAAGGTTTGGGCAAACTACTCTGACCTTTCAGCTAGTATCCCA... 1532
935 ArgAspAspValValValProCysGlyLysProValProSerSerIleAr 951
1533 CTGGATGCTGATAGGTTCTCTTGGACCAAGGTTTCATCTGCTGTGA 1582
951 gSerSer 953
1583 TGACACC 1589

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-328-111-216

seq_documentation_block:
; Sequence 216, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertli, Adnan

; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(595)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-216

alignment_scores:
Quality: 398.00 Length: 146
Ratio: 3.373 Gaps: 3
Percent Similarity: 80.822 Percent Identity: 56.164
alignment_block:
US-09-236-995D-2 x US-09-328-111-216/rev ..

Align seg 1/1 to reverse of: US-09-328-111-216 from: 1 to: 595
789 AsnThrHisAlaProThrHisLysAspTrpSerLeuGluLeuGluVal 805
445 TCTACCATGNTCCACACACAGCAGCTTATACCATGACCTTGTGAT 396
805 alPheSerLeuAspArgAspGlyGluLeuAsnLysTyrSerArgTyrLys 821
395 TTGTTTGAAGTCAGAGGATGTTGAG.....AAAGAAGCCTTCAGA 355
822 AsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAs 838
354 GAGGACCTTCATAACAGGATGCTTCTATGGCATGTTCCAGGATGAGTAA 305
838 nPheValGlyIleLeuSerGlnGlyLeuArgIleAlaProGluAlap 855
304 CTGGTGGGAATCTTGAGCCATGGGCTTCGAATGCCCACTGAAGCTC 255
855 roValThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuVal 871
254 CCATCAGAGTTACATGTTGGGAAAGGAATCTACTTTGCTGACATGCT 205
872 SerLysSerAlaClnTyrCysTyrValAspArgAsnAsnProValGlyLe 888
204 TCCAAGAGTGCATATCTGCTTTCCTCTCGCTTAAAGAAATACAGGACT 155
888 uMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysA 905
154 GCTGCTCTTATCAGAGGTAGCTCTAGGTCAGTGTAATGAACACTAGAGG 105
905 laThr...SerMetAspLysProProArgGlyLysHisSerThrLysGly 920
104 CCAATCCTAAGCCGAGGATGCTTCAAGGTAAACATAGCACCAAGGG 55
921 LeuGlyLysThrValProLeuGluSerGluPheVal 932
54 CTGGGCAAGATGGCTCCAGAGTCTGCCACTTCTGCCACTTCGTC 19

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-054-775C-1

seq_documentation_block:

```

; Sequence 1, Application US/09054775C
; Patent No. 6284504
; GENERAL INFORMATION:
; APPLICANT: Wei, Ying-Fei
; Yu, Guo-Liang
; Haseltine, William
; TITLE OF INVENTION: Human DNA Ligase III
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/054,775C
; APPLICATION NUMBER: 03-Apr-1998
; FILING DATE: 03-Apr-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/464,402
; FILING DATE: 05-JUN-1995
; APPLICATION NUMBER: PCT/US95/03939
; FILING DATE: 31-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Kenley K.
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PF161D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3417 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-054-775C-1

```

```

alignment_scores:
  Quality: 177.50      Length: 530
  Ratio: 0.752        Gaps: 24
  Percent Similarity: 44.528      Percent Identity: 20.377

```

```
alignment_block:
```

```
US-09-236-995D-2 x US-09-054-775C-1 ..
```

```
Align seg 1/1 to: US-09-054-775C-1 from: 1 to: 3417
```

```

11 GluTyrAlaLysSerGlyArgAlaSerCysLysSerCysArgSerProIl 27
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 GACTATGCCAAGCGTGCACAGCTGGCTGCAAAAATGCAAGGAAGAT 407
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 eAlaLysAspGlnLeuArgLeuGlyLysMetVal.....GlnAlaSerG 42
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 TGTGAAGGGCGTATGCCGAATTTGCAAGTGGTCCCAATCCCTTCTCAG 457
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 InPheAspGlyPheMetProMetTrpAsnHisAlaArgCysIlePheSer 58
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
458 AGTCTGGGGGTGATGAAGAGTGGTACCACATTAATGCATGTTTGAG 507
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 Lys.....LysAsnGlnIleLysSerValAspAs 68
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 AACTAGAGCGGGCGGGCCACCACAAAAAATCGAGGACCTCACAGA 557

```

```

68 pValGluGlyIleAspAlaLeuArgTrpAspAspGlnGluLysIleArgA 85
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 GCTGGAAGCGCTGGGAAGAGCTGGAAGATAATGAGAAGGAACAGATAACCC 607
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 snTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
608 AGCACATTGCAGATCTGCTCTTAAGGCAGCAGGTACA.....CCAAAG 651
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 GluLysCysThrIleGluIle..... 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 AAGAAAGCTGTTGTCACGGCTAAGTTGACACACCAGTCCAGGAGTACTTC 701
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 ..... 108
702 TCCAGTGAAGGCGCTCATTTGTCACCAAGTACCAATCCCGGAAATTTT 751
109 .....AlaProSerAlaArg 113
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
752 CTGGCTTTTCAGCCCAAGCCCAACAACTCTGGGGAAGCCCTCGAGCCCC 801
114 ThrSerCysArgArg, CysSerGluLysIleThrLysGlySerValArg. 129
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
802 ACCCTAAGAGAGTCTGCTTCAAGCAAAATGTACCCCGCAGCATAGA 851
130 LeuSerAlaLys..... 133
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
852 CTGCTGCTACGGGAGTTTCGAAAGTTATGCGCCATGGTGCCGATAATC 901
134 ....LeuGluSerGluGlyPro.....LysGlyIlePro..... 143
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
902 CTAGCTACACACAGAACAGCCAGATCATCCAGGAGCTCTCTTCGGAAAGC 951
144 .....Trp, TyrHisAlaAsnCysPhe..... 150
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
952 TCAGCAGGAGATGGTTTCCACGGTGATGTGTACCTAACAGTGAAGCTGCT 1001
151 .....PheGluValSerProSerAlaThrV 159
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1002 GCTGCCAGGAGTCATTAAAGACTGTTTACAACTTGAACGATAAGCAGATTG 1051
159 alGluLysPheSerGlyTrpAspThrLeuSerAspGluAspLysArgThr 175
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1052 TGAAGCTTTTCAGTCGCAATTTTAACTGCAACCCAGATGATATGCCACGG 1101
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 MetLeuAspLeuValLysLysAspValGlyAsnAsn..... 187
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1102 .....GACCTAGACGAGGTGACGTGTCAGAGACAATCAGAGTCTTCTT 1145
188 .GluGlnAsnLys.....GlySerLysArgLysLysSerGluA 200
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1146 TGAGCAGACGAAGTCTTTCCCGCCAGCTGCCAAGAGCCCTCTTACCATCC 1195
200 snAspIleAspSerTyrLysSerAlaArgLeuAspGluSerThrSerGlu 216
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1196 AGCAAGTGGATGAGTTC...CTTCTGGCGCTGTCCAAGCTCACCAGGAG 1242
217 GlyThrValArgAsnLysGlyGlnLeuValAspProArgGlySerAsnTh 233
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1243 GAT.....GAGCAGCAACAGGCCCTTACAGGACATTCCTCCAGGTG 1283
233 rSerSerAlaAspIleGlnLeuLysLeuLysGluGlnSerAspThrLeuT 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1284 TACAGCAATGACCTTAAATGCATCATCAGG..... 1314
250 rLysLeuLysAspGlyLeuLysThrHisValSerAlaAlaGluLeuArg 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1315 ..TTGATCAACATGATCTGAAGATCAACTCAGTGCACAAACATGTGTTA 1362
267 AspMetLeuGluAlaAsnGlyGlnAspThrSerGlyProGluArgHisLe 283
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1363 GACGCGCTTGACCCCAATGCCTATGAAGCTTCAAGGCTTCAAGGCTCGCGCACT 1412
283 uLeuasp..... 285

```



```
1413 GCAGGATGCTGGAGCGGCTCTTCACACGCGCAGGAGTGAGAGG 1452
286 .....ArgCysAlaAspGlyMetLeuPheGlyAlaLeuGlyPro 298
1463 AGCGGGCCACAGAGAGCTGTGAGCGTCCAGGCTCGCTGATGACACCT 1512
299 Cys...ProVal..... 301
1513 GTGAGCCCATGTTGGCGGAGGCGCTGCAAGTCCGTTGAGTATGCAATGAA 1562
302 ....CysAlaAsnGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerG 317
1563 GAAATGTCCTCAATGCGATGTC..... 1584
317 LysValSerGluTrpSerLysCysThrTyrSerAlaThrGluProVal 333
1585 .....TCTGAGATCAAGTACGATGGA...GAGCGAGTC 1614
334 ArgValLysLysLysTrpGlnIleProHisGlyThrLysAsnAspTyrLe 350
1615 CAGGTGCATAG.....AATGAGACCACTT 1640
350 uMetLysTrpPheLysSerGlnLysValLysLysProGluArgVal.... 365
1641 CAGTACTTCAGCGCGCAGTCTCAAGCGCTTCCTTCACAAAGGTGGCC 1690
366 .....LeuProMetSerProGluLysSerGlySerLys 377
1691 ACTTTAGGACTACATTCCTCCAGCGCTTTCCTGGGGCCACACATGATC 1740
378 AlaThrGlnArgThrSerLeuLeuSerSerLysGlyLeuAspLysLeuAr 394
1741 TTGGATTCTCAAGGCTTCTGATTGACACAAAGACAGGCAACCACTGCC 1790
394 gPheSerValValGlyGlnSerLysGluAlaAla 405
1791 CTTTGGGACTCTGGGAGTACACAAAGAGCAGCC 1824
```

seq_name: /cqn2_6/ptodata/2/ina/5B_COMB.seq:us-08-464-402-1

seq_documentation_block:

Sequence 1, Application US/08464402

Patent No. 5858705

GENERAL INFORMATION:

APPLICANT: WEI, ET AL.

TITLE OF INVENTION: Human DNA Ligase III

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSER: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/464,402

FILING DATE: June 5, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/03939

FILING DATE: 31 MAR 95

ATTORNEY/AGENT INFORMATION:

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 325800-388

TELECOMMUNICATION INFORMATION:

```
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3417 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-464-402-1

alignment_scores:
Quality: 173.50 Length: 530
Ratio: 0.735 Gaps: 24
Percent Similarity: 44.528 Percent Identity: 20.189

alignment_block:
US-09-236-995D-2 x US-08-464-402-1 ..
Align seg 1/1 to: US-08-464-402-1 from: 1 to: 3417
```

```
11 GluTyrAlaLysSerGlyArgAlaSerCysLysSerCysArgSerProI 27
358 GACTATGCCAAGCGTGGCACAGTGGCTGCAAAAAAATGCAAGGAAAGAT 407
27 eAlaLysAspGlnLeuArgLeuGlyLysMetVal.....GlnAlaSerG 42
408 TGTGAAGGCGGTATGCCGAATTTGGCAAGTGGTGGCCCAATCCCTTCTCAG 457
42 LnPheAspGlyPheMetProMetTrpAsnHisAlaArgCysIlePheSer 58
458 AGCTGGGGGTGATATGAAGAGTGGTACCACATTAATGTCATGTTTGAG 507
59 Lys.....LysAsnGlnIleLysSerValAspAs 68
508 AAATAGAGCGGGCGGGCCACCACAAAAAATCGAGGACCTCACAGA 557
68 pValGluGlyIleAspAlaLeuArgTrpAspAspGlnGluLysIleArgA 85
558 GCTGAAGGCTGGGAAGAGCTGGAAGATAATGAGAAGGAACAGATAACCC 607
85 snTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAlaProPro 101
608 AGCACATTGCACATCTGCTTCTTAAGGCAGCAGGTACA.....CCAAG 651
102 GluLysCysThrIleGluIle..... 108
652 AAGAAAGCTGTTGTCAGGCTAAGTTGACACCACCTGGCCAGGTGACTTC 701
108 ..... 108
702 TCCAGTGAAAGGCGCTCATTTGTCCACCACTACCAATCCCGGAAATTTT 751
109 .....AlaProSerAlaArg 113
752 CTGGCTTTTACCCCAAGCCCACTCTGGGAAGCCCTCGAGGCCCC 801
114 ThrSerCysArgArg.CysSerGluLysIleThrLysGlySerValArg. 129
802 ACCCCCTAAGAGAAGTCTGCTTCAAGCAAAATGTGACCCCGCAGGCATAAGGA 851
130 LeuSerAlaLys..... 133
852 CTGTCTGTACGGGAGTTTCGAAAGTTATGCGCCCATGCTGCCCGGATAATC 901
134 .....LeuGluSerGluGlyPro.....LysGlyIlePro..... 143
902 CTAGCTACAACACAGGAGCCAGATGACCCAGGACTTCCTTCGGAAAGGC 951
144 .....Trp.TyrHisAlaAsnCysPhe..... 150
952 TCAGCAGGAGATGTTTCCACGGTGATGTGTACCTAACACAGTGAAGCTGCT 1001
```



```

3390 ..ATTCGAGAACACAGAGATGTGGT..... 3413
808 LeuAspArgAspGlyGluLeuAsnLysTyr..... 817
3414 ...AATGCTGGCGCATCTTCACAGATACAATGCTTCGAATTCAAAA 3460
817 ..... 817
3461 AGTTGTCAACAAGAAGTTGAGGAGCGGTTCTGCCACCGACAGAAGGAAG 3510
818 ..SerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGly 833
3511 TGTCTGAGAGAAATCACACCACTACAATGACGAGCATGTTGTTTCATG 3560
834 SerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAl 850
3561 TCTCCTTTTCATTAAT.....GCCATTATTCATAAAGGGTTT..... 3596
850 aProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrP 867
3597 .GATGAGCGACATGATCATATAGGAGGAATGTTTGGGCCGGATTATT 3645
867 heAlaAspLeuValSerLysSerAlaGlnTyr..... 877
3646 TTGCTGAAACCTCTCAAAAGCACCAATATGTTTATGGAAATGGAGGA 3695
878 .....CysTyrValAspArgAs 883
3696 GGAACAGCGTGCCTTACACAAAGGACAGGTCATGCTATATATGTCACAG 3745
883 nAsnProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMetT 900
3746 ACA.....ATGCTCTTCTGTAGAGTACCCCTTGGGAAATCCT 3783
900 yGluLeuLysAlaThrSerMetAspLysProProArgGlyLysHis 916
3784 TTCTGCAGTTTAGCACCATGAATGCCCGCCCTCCAGGCCACAC 3833
917 SerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 933
3834 TCAGTC..... 3839
933 sTrpArgAspAspValValProCysGlyLysProValProSerSerI 950
3840 .....ATTGGTAGCCG.....AGCG 3855
950 leArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSer 966
3856 TCAATGG.....CTGGCATATGCTCAATATGTCATCTACAGAGGAGA 3899
967 GlnValLysMetGlnPheLeuLeu 974
3900 CAGGCATACCCAGATATCTATC 3923
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-387-7

```

```
3954 .GATGAGCGACATCATACATAGAGGAATGTTTGGGCGCGGATTTATT 4002
867 heAlaAspLeuValSerLysSerAlaGlnTyr..... 877
|||||:|||||:|||||:
4003 TTGCTGAAACTCTCAAAAGAACCAATATGTTTATGGAATTGGAGGA 4052
878 .....CysTyrValAspArgAs 883
|||||:|||||:|||||:
4053 GGAACAGCGTCCCTACACACAGGAGGTCATGCTATATATGTCACAG 4102
883 nasnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMet 900
|||||:|||||:|||||:|||||:
4103 ACA.....ATGCTCTTCTGTAGAGTGACCCCTGGGAAATCCT 4140
900 yrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysHis 916
|||||:|||||:|||||:|||||:
4141 TTCTGCAGTTTAGCACCATGAAATGGCCGCGCTCCAGGGCACCAC 4190
917 SerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 933
|||||:|||||:|||||:
4191 TCAGTC..... 4196
933 srPArgAspValValProCysGlyLysProValProSerSer 950
|||||:|||||:|||||:
4197 .....ATTGGTAGCCG.....AGCG 4212
950 leArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSer 966
|||||:|||||:|||||:|||||:
4213 TCAATGGG.....CTGGCATATGCTGAATATGTCATCATACAGAGGAGAA 4256
967 GlnValLysMetGlnPheLeu 974
|||||:|||||:|||||:
4257 CAGGCATACCCAGAGTATCTTATC 4280

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-387-9

seq_documentation_block:
; Sequence 9, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 9:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 4657 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 6...2855
US-09-196-387-9

alignment_scores:
Quality: 157.50 Length: 258
Ratio: 1.382 Gaps: 11
Percent Similarity: 44.186 Percent Identity: 22.868

alignment_block:
US-09-236-995D-2 x US-09-196-387-9 ..
Align seg 1/1 to: US-09-196-387-9 from: 1 to: 4657

763 TyrMetLysLeuHisCys.....AspIleThrProLeuAl 774
|||||:|||||:|||||:
3814 TATTGACTTTTCACTGTGTTAATCAGGGAACGATTTTGTGTCATCTGC 3863
774 aHisAspSerGluAspTyrLysLeuIleGluGlnTyrLeuLeuAsnThrH 791
|||||:|||||:|||||:
3864 TCCAGAGATAAAGATATCAGTCAGTGGGAAGAAGAGATGCAAAAGTACT. 3912
791 isAlaProThrHisLysAspTyrPheSerLeuGluLeuGluValPheSer 807
|||||:|||||:|||||:
3913 ..ATTGAGAACACAGAGATGGTGGT..... 3936
808 LeuAspArgAspGlyGluLeuAsnLysTyr..... 817
|||||:|||||:|||||:
3937 ...AATGCTGGCGCATCTTCAACAGATACAAATGCTCATTCGAATTCAAAA 3983
817 ..... 817
3984 AGTTGTCAACAAGAAAGTTGAGGAGCGGTTCTGCCACCACAGAAAGAG 4033
818 ..SerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGly 833
|||||:|||||:|||||:
4034 TGCTGAGGAGAAATCAACACCATCACAATGAGCCCACTGTTGTTTCATGGT 4083
834 SerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAl 850
|||||:|||||:|||||:
4084 TCTCTTTTCATTAA.....GCCATTATTCAATAAGGGTTT..... 4119
850 aProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrP 867
|||||:|||||:|||||:
4120 ..GATGAGCGACATGCATACATAGGAGGAATGTTTGGGCGCGGATTTATT 4168
867 heAlaAspLeuValSerLysSerAlaGlnTyr..... 877
|||||:|||||:|||||:
4169 TTGCTGAAAACTCCTCAAAAGAACCAATATGTTTATGGAATTGGAGGA 4218
878 .....CysTyrValAspArgAs 883
|||||:|||||:|||||:
4219 GGAACAGCGTCCCTACACACAGGAGGTCATGCTATATATGTCACAG 4268
883 nasnProValGlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMet 900
|||||:|||||:|||||:
4269 ACA.....ATGCTCTTCTGTAGAGTGACCCCTGGGAAATCCT 4306
900 yrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysHis 916
|||||:|||||:|||||:
4307 TTCTGCAGTTTAGCACCATGAAATGGCCGCGCTCCAGGGCACCAC 4356
917 SerThrLysGlyLeuGlyLysThrValProLeuGluSerGluPheVal 933
|||||:|||||:|||||:
```

4357 TCAGTC..... 4362
933 sTrpArgAspValValProCysGlyLysProValProSerSerI 950
4363ATTGGTAGCCG.....AGCG 4378
950 leArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSer 966
4379 TCAATGGG.....CTGGCATATGCTGAATATGTCTACAGAGAGAA 4422
967 GlnValLysMetGlnPheLeuLeu 974
4423 CAGGCATACCCAGATCTTATC 4446

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-07-861-458C-6

seq_documentation_block:

; Sequence 6, Application US/07861458C

; Patent No. 6232061

; GENERAL INFORMATION:

; APPLICANT: Marchionni, Mark Andrew

; APPLICANT: Johnson, Carl D.

; TITLE OF INVENTION: HOMOLOGY CLONING

; NUMBER OF SEQUENCES: 142

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 50z or 55SX

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/861,458C

; FILING DATE: 04/01/92

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 04585/014001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1638

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-07-861-458C-6

alignment_scores:

Quality: 121.50 Length: 370

Ratio: 0.774 Gaps: 16

Percent Similarity: 42.432 Percent Identity: 21.892

alignment_block:

US-09-236-995D-2 x US-07-861-458C-6 ..

Align seg 1/1 to: US-07-861-458C-6 from: 1 to: 1638

53 AlaArgCysIlePheSerLysLysAsnGlnIleLysSerValAspAspVa 69

|||||..... |||

706 GCGAAGCGCTCTCATTTGACCGCGCAAAATACCTTGTTCGTTCTTGATGATG 755

69 lCluGlyIleAspAlaLeuArgTrpAspAspGlnGluLysIleArgAsnTr 86
| |||..... |||

756 GTTCAAGAAGACACAATCCGTTGGGCTCAGGAACCTCGCCTTGCTGTT 805
|||..... |||

86 yValGlySerAlaSerAlaGlyThrSerThrAlaAlaProGlu 102
::: ::| |||..... |||..... |||.....

806 TGATCACCACAGAGATGTTGAAATTTCAAAV...GCACCTTACCAGAA 852
::: ::| |||..... |||..... |||.....

103 LysCysThrIleGluIleAlaProSerAlaArgThrSerCys..... 116
|||..... |||..... |||..... |||.....

853 TCCGAATTTATAGAAGTACACACCGTTGAAAAGTTATGAATGCTTCGAAT 902
|||..... |||..... |||..... |||.....

116 116

903 GTTGAATCATATGGCATGCGGTCGCTGCTATTGAGAGAGACGAGGATA 952
..... 116

953 TCTTACACAAAACCATTCATCTAACGAGCGGAAATCCAGCAGCTCTCATG 1002
.....

117ArgArgCysSerGluLysIleThrLysGlySerValArgLe 130
::: |||::| |||..... |||..... |||.....

1003 ATGATTTTCAAGTCATCGAACCGAAACATTCGAGAGAGATGGCCAGCT 1052
::: |||::| |||..... |||..... |||.....

130 uSerAlaLysLeuGluSerGluGlyProLysGlyIleProTrpTyrHisA 147
|..... |||..... |||..... |||..... |||.....

1053 GAACAGTAAACTGGAACTCGTGATTCATCTGCAANTGAATGATGATCACTC 1102
|..... |||..... |||..... |||..... |||.....

147 laAsnCysPhePheGluValSerProSer.....AlaThrValGluLys 161
|||::| |||..... |||..... |||.....

1103 CCTACTGTATAAGTCACTACTATAGTTCTCTCCAACGATGTGCGAA... 1149
|||::| |||..... |||..... |||.....

162 PheSerGlyTrpAspThrLeuSerAspGluAspLysArgThrMet..... 176
::: |||..... |||..... |||..... |||.....

1150GTTCTTTCGATGAAGATCGCAGCGCTTTGGCTTT 1184
.....

177LeuAspLeuValLysLys..... 182
::: |||..... |||..... |||..... |||.....

1185 TGCTGTTATCATGCCTCCAGGAATAGACATTCCTGCAAAATATGGTCTT 1234
.....

183AspValGlyAsnAsnGluGlnAsnLysGlySerLys 194
|||::| |||..... |||..... |||.....

1235 GCGTCATTCAGTGATATATGTTCCAACGAA..... 1266
|||::| |||..... |||..... |||.....

195 ArgLysLysSerGluAsnAspIleAspSerTyrLysSerAlaArgLeuAs 211
|||..... |||..... |||..... |||.....

1267GAAGATCAGTTAGATGACGAAAGTGGCGGATCGGTGAA 1304
|||..... |||..... |||..... |||.....

211 pGluSerThrSerGluGlyThrValArgAsnLysGlyGlnLeuValAsp 228
::: ::| |||..... |||..... |||.....

1305 GAGA.....TTAAGCAAAAGAGGAGCTCTGCTCAGTG 1336
::: ::| |||..... |||..... |||.....

228 roArgGlySerAsn.....ThrSerSerAlaAsp.....IleGlnLeu 240
::: ||| ||| |||..... |||.....

1337 GAAAACGGTCCCCCGTTTGGACCTATAAAATAGATCATGTCATTCATCTA 1386
|||..... |||..... |||..... |||.....

241 LysLeuLysGluGlnSerAspThrLeuTrpLysLeuLysAspGlyLeuLys 257
|||..... |||..... |||..... |||.....

1387 TTCTGAAGCACGCTGTTGATGTT.....CAGACGATGTC 1421
|||..... |||..... |||..... |||.....

257 stThrHisValSerAlaGluLeuArgAspMetLeuGluAlaAsnGlyC 274
::: ||| ||| |||..... |||.....

1422 AAATGGAATCTCGATCTTGAACACAGT...CTTCATGAGCTTGGAATA 1468
|||..... |||..... |||..... |||.....

274 lnAspThrSerGlyProGluArgHisLeu.....LeuAspArgCys 287
::: ||| |||..... |||..... |||.....

1469 ACAACACGGCTACACAGAAAGACATATGCCATCAAAATTCGTCGTACA 1518
|||..... |||..... |||..... |||.....

288 AlaAspGlyMetLeuPheGlyAlaLeuGlyProCysProValCysAlaAs 304
::: ||| ||| |||..... |||.....

1519 TCIGCTGGTGACATGTTT..... 1536
|||..... |||..... |||..... |||.....

```

304 nGlyMetTyrTyrAsnGlyGlnTyrGlnCysSerGlyAsnValSerG 321
1536 ..... 1536
321 luTrpSerLysCysThrTyrSerAlaThrGluProValArgValLysLys 337
1537 ..... 1537
338 LysTrpGlnIleProHisGlyThrLysAsnAspTyrLeuMetLysTrpPh 354
1582 TTCATGCAAAATC...CATGCGACCTTCTATGATTCCTCAAGAG...TT 1625
354 eLysSerGln 357
1626 CACATCCCAA 1635

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-685-576-2

```

seq_documentation block:
: Sequence 2, Application US/08685576
: Patent No. 5906819
: GENERAL INFORMATION:
: APPLICANT: kaibuchi, kozo
: APPLICANT: iwamatsu, akihiro
: APPLICANT: nakano, takashi
: APPLICANT: ito, masasaki
: TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/685,576
: FILING DATE: 24-JUL-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 7-325129
: FILING DATE: 20-NOV-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 8-17150
: FILING DATE: 05-JAN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 8-131206
: FILING DATE: 26-APR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Bent, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 16887/843
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5053 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Bovine
: FEATURE:
: NAME/KEY: CDS

```

```

; LOCATION: 1..4164
US-08-685-576-2

alignment_scores:
  Quality: 113.50      Length: 949
  Ratio: 0.268        Gaps: 42
  Percent Similarity: 44.573  Percent Identity: 18.230

alignment_block:
US-09-236-995D-2 x US-08-685-576-2 ..
Align seg 1/1 to: US-08-685-576-2 from: 1 to: 5053

20 CysLysSerCysArgSerProIleAlaLys.....As 30
1408 TGCAGTCTCTTAACTCGCTTAGAGAAAGTGCAAGAGGAGTTAGAA 1457
30 pGlnLeuArgLeuGlyLysMetValGlnAlaSer.....GlnPheAspG 45
1458 AGAGATTACCTTAAGGAAAAATGTGAATCAACATTAGACAATTAGAAA 1507
45 LysPheMetProMetTrpAsnHisAlaArgCysLlePheSerLysLys... 60
1508 GAGAAAAAGCAGCTTCTTCAGCACAAAAATGCAATATCAGCGGAAGCT 1557
61 .....AsnGlnIleLysSerVa 66
1558 GATCATGAACAGACACAAGAGCAAAATTTGGAGAAATGATGTTAACAGT 1607
66 lAspAspValGluGlyIleAspAlaLeuArgTrpAspAspGlnGluLysI 83
1608 AAAAGAT.....CAGCTTGAAGATTGAAAAA. 1635
83 leArgAsnTyrValGlySerAlaSerAlaGlyThrSerSerThrAlaAla 99
1636 ..AGAAATCAGAACTCTCAGATATCCACT..... 1662
100 ProProGluLysCysThrIleGluIleAlaProSerAlaArgThrSerCy 116
1663 .....GAGAAAGTGAATCAACTCCAGAGACAACTGGATGAAACCAATGC 1706
116 sArgArgCysSerGluLysIleThrLysGlySerValArgLeuSerAla 133
1707 TTTGCTCGCAACAGAAATCTGATACTCGAGCCCGGTTAAGG.....A 1747
133 ysLeuGluSerGluGlyProLysGlyIleProTrpTyrHisAla..... 147
1748 AACACAGGAGAGAAAGTTCAAAACAGATTTCAGCAGCTGGAATCTAACAA 1797
148 .....AsnCysPhePheGluValSerProSerAla 158
1798 AGAGATCTACAAGACAAAAATTCCTGCTGAGAGCTGCCCAAGTTAAACT 1847
158 rValGluLysPheSerGlyTrpAspThrLeuSerAspGluAspLysArg 175
1848 TGAAGAGAAATTTATCAATCTTCAGTCAGTCTTCTAGATCTGGAAGAGG. 1896
175 hrMetLeuAspLeuValLysLysAspValGlyAsnAsnGluGlnAsnLys 191
1897 .....GACCCGAAACCCAT 1908
192 GlySerLysArgLysLysSerGluAsnAspIleAspSerTyrLysSerAl 208
1909 GGATCA.....GAGATTATTAATGATTATCAAGGT...AGAATATC 1946
208 aArgLeuAspGluSerThrSerGluGlyThrVal..... 219
1947 TGGCTAGAGAGAGATGTAAGAAATGTAAATCTTTATTAGCAAAAGTAG 1996
220 ..ArgAsnLysGlyGlnLeuValAsp.....ProArgGly 230
1997 AGCTGGAGAGAGACAACTACAGGAGAGATTTACTGATTGGAAAAAGGAA 2046

```


440 snAlaGluValArgLysAlaArgArgLeuLysIleProIleValArgGlu 456
 1833 ATGTAAATAATGAGGC 1849
 457 GlyTyrIleGlyGluCysValLysArgThrLysCysCysHisLeuIleCys 473
 1850GGATTATCTATTTAAAAAACCAA 1873
 473 sIleAsnTrpAsnAlaLeuGluSerSerLysGly***ThrValThrVal 490
 1874AAAAAGAAATAAAGAAAGAGGTGGAATAATACGTCTGAAA 1910
 490 ysValLysGlyArgSerAlaCysSer***SerPro***ValCysLysAsn 506
 1911 AGGAGCCTGATGAATCCAAAGACACATCAATCCTTTTACTATTGG 1960
 507 ThrAlaHisIlePro*****TrpGluLysHisIleGlnCys 520
 1961 GTTGACATATGTAAAGATTCATACATTCGGAACAAATAAATGT.... 2056
 520 s***LeuLysHisValLeuThr***His***ValCysThrG 534
 2011 ATGTTTACAAATGGTAACAGATAAATCTGGAACAAATAAATGT.... 2056
 534 lyTyrTyrValLeuGlnIleLeuGlnAspAspGlySerGluCysTyr 550
 2057ATAATGATTGTGAA 2071
 551 ValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLys 567
 2072 TGTTTTAAAGATGGATTACACAAAA 2097
 567 sLeuGluGluMetSerLysThrGluAlaIleLysGluPheLysArgLeuP 584
 2098 AAAAGACGAATGGGGAAAATAGTACACATTTTAAACACGAAA 2141
 584 heLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAsn 600
 2142 ATATTAAGAGTAGAGGAGGTAGT 2164
 601 PheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrGlyVa 617
 2165GACATACGGCAGAAATTAATCCCATTTGATCAGGATTATGTCT 2208
 617 lLysLysAlaProLysArgLysAspIles 627
 2209 TCAATACAAATTTGCAAGAAAGAAATTTTGAAGGCGATTCGGAACGCTT 2258
 627 erGluMetLysSerSerLeuAlaProGlnLeuLeuLeuMet 641
 2259 CGGAAGAAAATCCGAAATAGTCTGGATGCGAGAGGAGGAGGAACTA 2308
 642 LysMetLeuPheAsnValGluThrTyrArgAlaAlaMetMetGluPheG 658
 2309 AAACACCTTCGCAAAATCATTGAAGGTGA 2337
 658 u***AsnMetSerGluMetProLeuGly 667
 2338 AGACAAATAATCAAGAGCATCTGTGTGT 2365

seq_name: /cgn2_6/ptodata/2/ina/5B_comb.seq:US-08-487-826B-13

seq documentation block:

; Sequence 13 Application US/08487826B
 ; Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

181 yLysAspValGlyAsnAsnGluGlnAsnLysGlySerLysArgLysLys 197
 1182 AAAAATATGAAATGGAGCATCAGCTGTAGTAGCAAAAACGG 1225
 198SerGluAsnAspIleAspSerTyrLysSerAlaAr 209
 1226 GATGAGGTGGTACAACTACTACTAATTTATGATGATATGAAAAAAT 1275
 209 gLeuAspGluSerThrSerGluGlyThrValArgAsnLysGlyGlnLeuV 226
 1276 TTATGACGAACTAATAAAGTGAATATAGAACCG 1310
 226 alAspProArgGlySerAsnThrSerSerAlaAspIleGlnLeuLysLeu 242
 1311 TTGATAAATTTTGGAAAAATTAAGTAATGAAGAAATATGCACAAAAGTT 1360
 243 LysGluGlnSerAspThrLeuTrpLysLeuLysAspGlyLeuLysTh 258
 1361 AAAGACGAAGAGGAGGAAACAATT 1384
 258 rHisValSerAlaAlaGluLeuArgAspMetLeuGluAlaAsnGlyGlnA 275
 1385GATTTTAAAAACGTTAATAGTGATA 1409
 275 spThrSerGlyProGluArgHisLeuLeuAspArgCysAlaAspGlyMet 291
 1410 GTACTAGTGTGCTAGTGGCACT 1432
 292 LeuPheGlyAlaLeuGlyProCys 299
 1433 AATGTTGAAGTCAGAGACATTTATCTGTTCAAAATATGCGCAACCTGT 1482
 299 sProValCysAlaAsnGlyMetTyrTyrTyrAsnGlyGlnTyrGlnCysS 316
 1483 CCCTTATGTGGAGTGAAGGAGGAAATAATG 1514
 316 erGlyAsnValSerGluTrpSerLysCysThrTyr 327
 1515 GTGGTAGTAGTATGATGGAAGAGAGAAAATAATGGCAAGTGC 1558
 328 SerAlaThrGluProValArgValLysLysLysTrpGlnIleProHisG 344
 1558 1558
 344 yThrLysAsnAspTyrLeuMetLysTrpPheLysSerGlnLysValLysL 361
 1559AAGAGTGAACACTTATG 1577
 361 ysProGluArgValLeuProProMetSerProGluLysSerGlySerLys 377
 1578 AGCCTAAACCGACAAAGAGGT 1600
 378 AlaThrGlnArgThrSerLeuLeuSerSerLysGlyLeuAspLysLeuAr 394
 1601ACTACTATTCAATCCCTTAAAGAGTGGTAAAGGACATGATGATGCA 1647
 394 gPheSerVal 397
 1648 AGAAAAATTAACAAATTTTGTGATGAAAAAATGGTGATACAAATAATA 1697
 398ValGlyGlnSerLysGluAlaAlaAsn 406
 1698 GTGGTGTAGTGTGACGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1747
 407 GluTrpIleGluLysLeuLysLeuAlaGlyAlaAsnPheTyrAlaArgVa 423
 1748 GAATTGTGTAAGAAATGGAATGTTTAAAGGTGCAAGATCTAGTGAAGT 1797
 423 lValLysAspIleAspCysLeuIleAlaCysGlyGluLeuAspAsnGluA 440
 1798 TGGACACGATGAGGATGACGAGGAGGATTTATGAAA 1832


```

520 s***LeuLysHis.....ValLeuThr***His***ValCysThrG 534
|||||:|||||:|||||
9331 ATGTTTACAAAATGTAACAGATAAATGTGGAACAAATAAATGT.... 9376
|||||:|||||:|||||
534 LyTyrTyrValLeuGlnIleIleGluGlnAspGlySerGluCysTyr 550
|||||:|||||:|||||
9377 .....ANTAATGATTGTCAA 9391
|||||:|||||:|||||
551 ValPheArgLysTrpGlyArgValGlySerGluLysIleGlyGlnLy 567
|||||:|||||:|||||
9392 TCGTTTAAAGAGATGATTACA.....CAAAA 9417
|||||:|||||:|||||
567 sleuGluGluMetSerLysThrGluAlaIleLysGluPheLysArgLeuP 584
|||||:|||||:|||||
9418 AAAAGACGAATGGGGGAAA.....ATAGTACACATTTTAAACGCNAA 9461
|||||:|||||:|||||
584 heLeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAsn 600
|||||:|||||:|||||
9462 ATATTAAAGGTAGAGGAGTAGT..... 9484
|||||:|||||:|||||
601 PheArgLysGlnProGlyArgPheTyrProLeuAspValAspTyrGlyVa 617
|||||:|||||:|||||
9485 .....GACATACGCGCAGAAATTAATCCCATTTGATCAGGATTATGTCT 9528
|||||:|||||:|||||
617 lLysLysAlaProLysArg.....LysAspIles 627
|||||:|||||:|||||
9529 TCATATACAAATTGCAAGAGAGAAATTTTGAAGGCGATTCCGAAGACGCTT 9578
|||||:|||||:|||||
627 erGluMetLysSer.....SerLeuAlaProGlnLeuGluLeuMet 641
|||||:|||||:|||||
9579 CCGAAGAAAAATCCGAAAAATAGTCTGGATGCGAGAGGAGGAGGAACTA 9628
|||||:|||||:|||||
642 LysMetLeuPheAsnValGluThrTyrArgAlaAlaMetMetGluPheGl 658
|||||:|||||:|||||
9629 AAACACCTTCGCGAAATC.....ATTGAAAGTGA 9657
|||||:|||||:|||||
658 u***AsnMetSerGluMetProLeuGly 667
|||||:|||||:|||||
9658 AGACAATAATCAAGAGCATCTGTTGGT 9685

```


568 euGlucMetSerLysThrGluAlaLeuLysGluPheLysArgLeuPhe 584
|||||
118 TGGAGGAGATGTCACAAACTGAGGCAATCAAGGAATTCACAAAGATATTT 167
|||||
585 LeuGluLysThrGlyAsnSerTrpGluAlaTrpGluCysLysThrAsnPh 601
|||||
168 CTTGAGAGACTGGAACTCACTGGGAAGCTGGGAATGTAACCAATTT 217
|||||
601 eArgLysGlnProGlyArgPheTyPrProLeuAspValAspTyPrGlyValL 618
|||||
218 TCGGAGACGCTGGGAGATTTTACCACCTTGATGTTGATGATGTTGTTA 267
|||||
618 yLysAlaProLysArgLysAspLeSerGluMetLysSerSerLeuAla 634
|||||
268 AGAAGACACCAAGGAAAGATATCAGTGAATGAAAGTCTCTTGCT 317
|||||
635 ProGlnLeuLeuLeuMetLysMetLeuPheAsnValGluThrTyPrAr 651
|||||
318 CCTCAATGCTAGACTGAGATCTGAGGATTTTCCACCTTGATGTTGATGTTGTTA 367
|||||
651 gAlaAlaMetMetGluPheGlu***AsnMetSerGluMetProLeuGlyL 668
|||||
368 AGCTGCTATGATGGAATTTGAAATTAATATGTCAGAAATGCTCTTGGA 417
|||||
668 yLysLeuSer*****AsnLeuGlu***GlyPheGluAlaLeuThr***** 684
|||||
418 AGCTAAGCAAGGAAATTTGAGAAAGGATTTGAGCATTAACACTGAGATA 467
|||||
685 *****LeuPheGluGlyHisArg***SerSerThrGlyLeu***GluL 701
|||||
468 CAGAAATTTATT.GAAGCACACCGCTCATCAAGCACTGGCTG.TTAGAGAA 515
|||||
701 sAla***LeuLeu*****PheSerLeuLeuSerLeuLeuP 718
|||||
516 AGCTTAATGTTGCTCGAGCAATCGCTTTTCACTCTTATCCCTCTAT 565
|||||
718 heileLeuLeuTyPrGlyMetArgMetIleSerTyPrSerLysAlaLys 734
|||||
566 TATN.CTCATATTATACGGGATGAGGATGATTG.ATGATCAAGCGGAA 613
|||||
735 MetLeuGluAlaLeuGlnAspIleGluLeuAlaSerLys 747
|||||
614 ATGCTTGAAGCTCTGCAGGATATTGAATTTGCTTCAAG 652
|||||
seq_name: gb_est1:AV833309

seq_documentation_block:
LOCUS AV833309 714 bp mRNA EST
DEFINITION AV833309 K. Sato unpublished cDNA library: Hordeum vulgare subsp.
vulgare shoots germination Hordeum vulgare subsp. vulgare cDNA
clone bags5K04, mRNA sequence.
ACCESSION AV833309
VERSION AV833309.1 GI:14525398
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare.
ORGANISM Hordeum vulgare subsp. vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
Sato,K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2001)
Contact: Kazuhiro Sato
Research Institute for Bioresources
Okayama University, Barley Germplasm Center
Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
Email: kzsato@rib.okayama-u.ac.jp,
URL:http://www.rib.okayama-u.ac.jp/barley/
Sato,K., Saisho,D., Takeda,K., Shini,T. and Kohara,Y. Direct
submission:
database:http://www.shigen.nig.ac.jp/barley/Barley.html.

FEATURES
Source Location/Qualifiers
1..714
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna NiJo"
/db_xref="taxon:112509"
/clone="bags5K04"
/clone_lib="K. Sato unpublished cDNA library: Hordeum
vulgare subsp. vulgare shoots germination"
/tissue_type="shoots"
/dev_stage="germination"
BASE COUNT 249 a 121 c 162 g 182 t
ORIGIN

alignment_scores:
Quality: 783.00 Length: 240
Ratio: 4.015 Gaps: 4
Percent similarity: 81.250 Percent identity: 71.667

alignment_block:
US-09-236-995D-2 x AV833309 ..

Align seg 1/1 to: AV833309 from: 1 to: 714

509 HisIlePro*****TrpGluLysHisIleGlnCys***LeuLysHisVa 525
|||||
6 CACGAGGCTGGAAGATGCGCAAGCAATTTACAAATACAAACCTTAACAT.. 53
|||||
525 lLeuThr***His***ValCys.ThrGlyTyPrTyPrValLeuGlnIle 541
|||||
54 .TTCAGCATGACACAAAGGTGTTAACAGCTACTATATATCTTACATCATC 102
|||||
542 GluGlnAspAspGlySerGluCysTyPrValPheArgLysTrpGlyArgVa 558
|||||
103 GAAGAGGATGATGGAGTGAATGCTATGTTTTCGAAAGTGGGGCGAGT 152
|||||
558 lGlySerGluLysIleGlyGlnLysLeuGluGluMetSerLysThrG 575
|||||
153 TGGCAGTGAAGGATTTGTTGAAAGAACTGGAGGAGATGTCAAAACCTG 202
|||||
575 luAlaIleLysGluPheLysArgLeuPheLysThrGlyAsnSer 591
|||||
203 ACCTAATACATGAATTTTAAAGATTTTTCGAAAGAGACTGGAAACCCC 252
|||||
592 TrpGluAlaTrpGluCysLysThrAsnPheArgLysGlnProGlyArgPh 608
|||||
253 TGGGAAGCATGGGAACAAACAAATTTTCAGAAAGCAGCTGGGAGATT 302
|||||
608 eTyPrProLeuAspValAspTyPrGlyValLysLysAlaProLysArgLysA 625
|||||
303 TTATCCACTTGACATTTGATTACGGAGTTAAGCAAGCACCAGCAAGGAAAG 352
|||||
625 spIleSerGluMetLysSerSerLeuAlaProGlnLeuLeuGluLeuMet 641
|||||
353 ACATCAGCAAAATGAAAGATTCACCTGCTCCTCAGGTGCTGGAACCTCATG 402
|||||
642 LysMetLeuPheAsnValGluThrTyPrArgAlaAlaMetMetGluPheGl 658
|||||
403 ATGATGCTTTTCAATGTTTGAACATATATAGGCTGCTATGATGGAATTTGA 452
|||||
658 u***AsnMetSerGluMetProLeuGlyLysLeuSer*****AsnIleG 675
|||||
453 AATCAATATGCGCAAAATGCCCTTGGGAAATTAAGCAAGGAAATATATCC 502
|||||
675 lu***GlyPheGluAlaLeuThr*****LeuPheGluGlyHis 691
|||||
503 AGAAGAGGATTTGAACCATTAATGAGATACAAATCTACTGAT...GAC 549
|||||
692 Arg***Ser.SerThrGlyLeu***GluLysAla***LeuLeu***** 708
|||||
550 ACTGCAATCAAGAACTGCTCTTAGAGAGAGCTTGTATTGTTGTGCAAG 599
|||||
708 *****PheSerLeuLeuSerLeuPheIleLeuLeuTyPrGly 724

600 CAATCGTTCTTCACTCTTATTCCTTCTGTCATCCACATATTATACATG 649
 725 MetArgMetIleSerTyrSerIleAlaLysMetLeuCluAlaLeuGlnAs 741
 650 ATAAGGATGACCTG.ACAATGAAGCGAAATGCTTGAAGCTCTTCAGGA 698
 741 pileGluileAlaSer 746
 699 TATTGAAATTGCTTCT 714

seq_name: gb_est2:BF255013

seq_documentation_block:

LOCUS BF255013 607 bp mRNA EST 23-FEB-2001
 DEFINITION HVSMEF0005L18f Hordeum vulgare seedling root EST library HVCDNA0007
 (etiolated and unressed) Hordeum vulgare cdna clone
 HVSMEF0005L18f, mRNA sequence.

ACCESSION BF255013

VERSION BF255013.2 GI:13117445

KEYWORDS EST.

SOURCE barley.

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.

REFERENCE

AUTHORS

Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
 , Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
 , T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
 Wood, T.

Development of a genetically and physically anchored EST resource

for barley genomics

Unpublished (2000)

On Nov 16, 2000 this sequence version replaced gi:11184130.

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: AATTACCTCTACTAAAGGG

High quality sequence stop: 551.

FEATURES

source

1..607

/organism="Hordeum vulgare"

/cultivar="Morex"

/db_xref="taxon:4513"

/clone="HVSMEF0005L18f"

/clone_lib="Hordeum vulgare seedling root EST library

HVCDNA0007 (etiolated and unressed)"

/tissue_type="Seedling root"

/lab_host="TJCL21"

/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For

more details on library preparation and sequence analysis

see http://www.genome.clemson.edu/projects/barley/ro

order a clone see http://www.genome.clemson.edu/orders"

165 a 113 c 163 g 166 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 714.00 Length: 146

Ratio: 4.993 Gaps: 0

Percent Similarity: 97.945 Percent Identity: 91.781

alignment_block:

US-09-236-995D-2 x BF255013

Align seg 1/1 to: BF255013 from: 1 to: 607

837 ThrAsnPhaValGlyIleLeuSerGlnGlyLeuArgIleAlaProGln 853

1 ACAAATTTTATTGGATCTTCTAGTCAAGACTAAGATAGACCTCCTGA 50
 853 uAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaA9pL 870
 51 GGCACCCGTGAGGGGCTATATGTTGGCAAGGCCCTCTACTTTGCAGATT 100
 870 euValSerLysSerAlaGlnTyrCysTyrValAspArgAsnAsnProVal 886
 101 TAGTAGCAAGAGTGCACAGTATTGTTATGTGATAGAGAAATAATCCGACT 150
 887 GlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuL 903
 151 GCGTTGATGCTTCTTCAGGTTGCTTAGAGACATGCATGAACCTGAA 200
 903 sLysAlaThrSerMetAspLysProProArgGlyLysHisSerThrLysG 920
 201 AAAGCAACGGCAATGGACAAACCTCCAGAGAGGAGCATTCGACCAAG 250
 920 LyLeuGlyLysThrValProLeuGluSerGluPheValLysTrpArgAsp 936
 251 GCTTAGCAAAACTGTGCCACTAGAGTCGGAGTTGTTAAATGGAGGAT 300
 937 AspValValProCysGlyLysProValProSerSerIleArgSerSe 953
 301 GATGTCGTGCTGCTTGTGGCAAGCCAGTCGACGATCTATCAGGGCATC 350
 953 xCluLeuMetTyrAsnGluTyrIleValTyrAsnThrSerGlnValLysM 970
 351 TGAGTCTCTGACACAGAGTATATAGTGTACAACACAGCTCAGGTGAAGA 400
 970 etGlnPheLeuLysValArgPheHisLysLysArg 982
 401 TGCAGTCTCTGTTGAAGGTCAAAATCCCTCAAGCGT 438

seq_name: gb_est2:BG280821

seq_documentation_block:

LOCUS BG280821 861 bp mRNA EST 21-FEB-2001
 DEFINITION 602401166F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4543433 5',
 mRNA sequence.

ACCESSION BG280821

VERSION BG280821.1 GI:13029745

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTDP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCMI225 row: c column: 18

High quality sequence stop: 825.

FEATURES

source

1..861

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4543433"

/clone_lib="NIH_MGC_20"

/tissue_type="melanotic melanoma"

/lab_host="PH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dr priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

/clone_lib="k. Sato unpublished cDNA library: Hordeum
vulgare subsp. vulgare shoots germination"
/tissue_type="shoots"
/dev_stage="germination"
180 a 181 c 132 g 168 t

BASE COUNT
ORIGIN

alignment_scores:
Quality: 642.00 Length: 136
Ratio: 4.827 Gaps: 1
Percent Similarity: 97.794 Percent Identity: 90.441

alignment_block:

US-09-236-995D-2 x AV834168/rev ..

Align seg 1/1 to reverse of: AV834168 from: 1 to: 661

```

848 ArgfileAlaProGluAlaProValThr.GlyTyrMetPheGlyLysG 864
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
661 AGGATACACCTCTTGAGGACCCGCGGCTATATGTTTGGCAAG 612
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
864 lyLeuTyrPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrVal 880
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
611 GCCTCTACTTTGCAGATTTAGTAAGCAAGAGTGCACAGTATTGTTATGTG 562
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
881 AspArgAsnProValGlyLeuMetLeuSerGluValAlaLeuG1 897
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
561 CATAGAAAATCCGACTGGCTGTGATGCTCTTCTGAGGTTGCTCTAGG 512
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
897 yAspMetTyrGluLeuLysLysAlaThrSerMetAspLysProProArg 914
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
511 AGACATTCATGACTGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 462
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
914 lyLysHisSerThrLysGlyLeuGlyLysThrValProLeuGluSerGlu 930
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
461 GGAAGCATTCGACCAAGGCTTAGCAAAACTGTGCCACTAGAGTCGGAG 412
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
931 PheValLysTyrArgAspValValValProCysGlyLysProValPr 947
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
411 TTGTGTAATGAGGAGTATGTCGTCGCTGTCGTCGTCGTCGTCGTCG 362
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
947 oSerSerLeuArgSerSerGluLeuMetTyrAsnGluTyrValTyrA 964
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
361 AGCATCTATCAGGCGATCTGAGCTCTCTGACACGAGTATATATGTACA 312
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
964 snThrSerGlnValLysMetGlnPheLeuLeuLysValArgPheHisHis 980
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
311 ACACAGCTCAGTGAAGTGCAGTCTTGTGAAAGGTCAAAATTCGGTCAC 262
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
981 LysArg 982
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
261 AAGCGT 256

```

seq_name: gb_est2:BG031594

seq_documentation_block:

LOCUS BG031594 990 bp mRNA EST 24-JAN-2001
DEFINITION 602299739F1 NIH_MGC-87 Homo sapiens cDNA clone IMAGE:4394288 5',
mRNA sequence.

ACCESSION BG031594

VERSION BG031594.1 GI:12422028

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 990)

NIH-MGC <http://mhc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM10089 row: m column: 09

High quality sequence stop: 678.

FEATURES

source

Location/Qualifiers

1..990

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4394288"

/clone_lib="NIH_MGC_87"

/tissue_type="mammary adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: breast; Vector: pCMV-SPORT6; Site: 1; NotI;

Site: 2; SalI; Cloned unidirectionally; oligo-gt primed.

Average insert size 1.383 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

BASE COUNT 274 a 218 c 251 g 247 t

ORIGIN

alignment_scores:

Quality: 642.00 Length: 232

Ratio: 3.326 Gaps: 2

Percent Similarity: 83.190 Percent Identity: 54.310

alignment_block:

US-09-236-995D-2 x BG031594

Align seg 1/1 to: BG031594 from: 1 to: 990

```

750 GlyPheAspSerAspSerAspGluSerLeuAspLysTyrMetLysLe 766
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
9 GGGTCTGATGATAGCAGCAGGATCCCATCGATGTCAACTATGAGAAGCT 58
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
766 whiLysCysAspIleThrProLeuAlaHisAspSerGluAspTyrLysLeuI 783
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
59 CAAACTGACATTAAAGTGGTGTGACAGAGATTCTCAAGAGCCGAGATCA 108
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
783 leGluGlnTyrLeuLeuAsnThrHisAlaProThrHisLysAspTyrSer 799
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
109 TCAGGAAGTATGTTAAGAACACTCATGCAACACACAAATGCGTATGAC 158
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
800 LeuGluLeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLy 816
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
159 TTGGAAGTCATCGATATCTTTAAGATAGAGCGTGAAGCGAATGCCAGCG 208
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
816 sTyrSerArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisG 833
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
209 TTACAAAGCCCTTTAAG...CAGCTTCATAACCCGAAGATTGCTGTGGCAG 255
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
833 lySerArgLeuThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIle 849
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
256 GGTCCAGGACCAACCACTTGTCTGGGATCTGTCCAGGGTCTTCGGATA 305
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
850 AlaProProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTy 866
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
306 GCCCGCGCTGAAGCGCCCGTGACAGGCTACATGTTGGTAAAGGATCTTA 355
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
866 rPheAlaAspLeuValSerLysSerAlaGlnTyrCysTyrValAspArgA 883
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
356 TTTTCGCTGACATGCTCTCCAGAGAGTCCCAACTACTGCCCATACGCTC 405
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
883 snAsnProValGlyLeuMetLeuSerGluValAlaLeuGlyAspMet 899
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
406 GAGACCAATAGGCTTAATCTCTGTGGGAACTTGCCTTGGAAACATG 455
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
900 TyrGluLeuLysLysAlaThrSerMetAspLysProProArgGlyLysH 916
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

456 TATGAACGAGCAGCTTCACATATCAGCAAGTTACCAAGGCGCAAGCA 505
916 sSerThrTyrGlyLeuGlyLysThrValProLeuGluSerGluPheValL 933
506 CAGTGTCAAAGGTTGGGCAAACTACCCCTGATCCCTTCAGCTAACATTA 555
933 yStrpArgAspValValValProCysGlyLysProValProSerSer 949
556 GT...CTGATGGTGTAGACGTTCTCTTGGGACCGGATTCATCTGGT 602
950 IleArgSerSerGluLeuMetTyrAsnGluTyrIleValTyrAsnThrSe 966
603 GTCAATGACACCTCTCTACTATAACAGGTACATGTCTATGATATGTC 652
966 rGlnValLysMetGlnPheLeuLeuLysValArgPheHisLys 981
653 TCAGTA.CATCTGAAGTATCTCTGAAACTTGAATTCATTAATGTAAG 697

seq_name: gb_est2:BI091452
seq_documentation_block: 834 bp mRNA EST 20-JUN-2001
LOCUS BI091452 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5000263 5',
DEFINITION mRNA sequence.
ACCESSION BI091452
VERSION BI091452.1 GI:14509782
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1031 row: n column: 08
High quality sequence stop: 804.

FEATURES
SOURCE
1..834
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5000263"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 230 a 206 c 214 g 184 t
ORIGIN

```

```

alignment_scores:
  Quality: 618.00
  Ratio: 2.971
  Percent Similarity: 74.820
  Percent Identity: 47.482

alignment_block:
US-09-236-995D-2 x BI091452
Align seg 1/1 to: BI091452 from: 1 to: 834

seq_name: gb_est2:BG429402
seq_documentation_block: 632 bp mRNA EST 14-MAR-2001
LOCUS BG429402 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4613500 5',
DEFINITION 602499622F1

```

mRNA sequence.
 BG429402
 VERSION BG429402.1 GI:13335908
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: NIH.gov
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLCMI363 row: k column: 05
 High quality sequence stop: 628.
 Location/Qualifiers
 1..632
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4613500"
 /clone_lib="NIH_MGC.75"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="organ: kidney; Vector: pDNR-Lib (Clontech); Site_1: SfiI (ggccctcgcc); Site_2: SfiI (ggccattatgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
 BASE COUNT 179 a 146 c 151 g 156 t
 ORIGIN
 alignment_scores
 Quality: 608.00 Length: 210
 Ratio: 3.514 Gaps: 2
 Percent Similarity: 82.381 Percent Identity: 55.714
 alignment_block
 US-09-236-995d-2 x BG429402 ..
 Align seg 1/1 to: BG429402 from: 1 to: 632
 759 LeuAspLysTyrMetLysLeuHisCysAspLeuThrProLeuAlaHi 775
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 6 ATCGATGTCACATCTGAGAGCTCAAACTGACATTAAGGTGGTGACAG 55
 775 sAspSerGluAspTyrLysLeuIleGluGlnTyrLeuLeuAsnThrHis 792
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 56 AGATTCTGAAGAAGCGGAGATCATCAGGAAGTATGTTAAGAACATCATG 105
 792 laProThrHisLysAspTrpSerLeuGluLeuGluValPheSerLeu 808
 || ||||| :||||| :||||| :||||| :||||| :||||| :||
 106 CAACACACACAAATCGGTGATGATCGGAAGTCATCGATATCTTTAAGATA 155
 809 AspArgAspGlyGluLeuAsnLysTyrSerArgTyrLysAsnLeuHi 825
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 156 GAGCGTGAAGGCGATGCCAGGCTTACAGCCCTTAAG...CACCTTCA 202
 825 sAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPheValGlyI 842
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 203 TAACCGAAGATTGCTGTCGCGGGTCCAGGACCAACCACTTGTGCGGA 252

842 leuSerGlnGlyLeuArgIleAlaProGluAlaProValThrGly 858
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 253 TCTCTGCCAGGGTCTTCGATAGACCCCGCCGTAAGCGCCGTCGACAGGC 302
 859 TyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerLysSerAl 875
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 303 TACATGTTGGTAAAGGGATCTATTTCGCTGACATGGTCTCCAAGAGTGC 352
 875 aGlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMetLeuLeu 892
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 353 CAACACTCTGCCATAGCTCTCAGGAGACCAATAGCTTAATCCTGTTGG 402
 892 exGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaThrSerMet 908
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 403 GAGAAGTTCCTTGGAAACATGTATGACTGAACGACGCTTCACATATC 452
 909 AspLysProProArgGlyLysHisSerThrLysGlyLeuGlyLysThrVa 925
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 453 AGCAAGTTACCAAGGCAAGCACAGTGTCAAGGTTTGGGCAAACTAC 502
 925 lProLeuGluSerGluPheValLysTrpArgAspValValValProC 942
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 503 CCTGATCCTTCAGCTAACATTAGT...CTGGATGGTGTAGACGTTCTCTC 549
 942 ysGlyLysProValProSerIleArgSerSerGluLeuMetTyrAsn 958
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 550 TTGGGACCGGGATTCATCTGGTGTGAATGACACCTCTACTATATAAAC 599
 959 GluTyrIleValTyrAsnThrSerGlnVal 968
 :||||| :||||| :||||| :||||| :||||| :||||| :||
 600 GAGTACATTGTCTATGATTGCTCAGGTA 629
 seq_name: gb_est2:BE941860
 seq_documentation_block:
 LOCUS BE941860 419 bp mRNA EST 03-OCT-2000
 DEFINITION EST421439 MGHG Medicago truncatula cDNA clone pMGHG-6B18, mRNA sequence.
 ACCESSION BE941860
 VERSION BE941860
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
 REFERENCE 1 (bases 1 to 419)
 AUTHORS Cote, F., Ojanen-Reuths, T., Hahn, M.G., VandenBosch, K., Hur, J., Beremand, P., Endre, G., Town, C.D., Bowman, C.L., Craven, M.B. and Cho, J., Fraser, C.M.
 TITLE ESTs from seedling roots of Medicago truncatula after treatment with beta glucan elicitor preparation from Phytophthora sojae
 JOURNAL Unpublished (2000)
 COMMENT Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu
 University of Georgia name: G269712e TIGR sequence name: MTUAP09TK
 More information is available at: <http://chryslie.tamu.edu/medicago>
 Seq primer: SKmod (CTA gaa CTA gtg gAT CC).
 FEATURES
 Location/Qualifiers
 1..419
 /organism="Medicago truncatula"
 /cultivar="Al7"
 /db_xref="taxon:3880"
 /clone="pMGHG-6B18"
 /clone_lib="MGHG"
 /tissue_type="Roots from four day old seedlings"

/dev_stage="2 days after treatment with beta glucan
elicitator preparation from Phytophthora sojae"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."
BASE COUNT 131 a 87 c 101 g 100 t
ORIGIN

alignment_scores: Quality: 589.00 Length: 139
Ratio: 4.496 Gaps: 0
Percent Similarity: 94.245 Percent Identity: 76.978
alignment_block:
US-09-236-995D-2 x BE941860 ..
Align seg 1/1 to: BE941860 from: 1 to: 419

790 ThrHisAlaProThrHisLysAspTTPSerLeuGluLeuGluValPh 806
|||||
1 ACTCATGCCCCACATAGGAGTCTTGAGCTAGAGAGAAGT 50
806 eSerLeuAspArgAspGlyLeuLeuAsnLysTyrSerArgTyrLysAsnA 823
|||||
51 AGCACTAGAAGAGAGAGGTGAATGAATCAATATGCCGTTACAGAGACA 100
823 snLeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeuThrAsnPhe 839
|||||
101 AAGTAGGCAACAGAAATGCTCTATGGCATGGTCTTAGGCTGACGAACTTT 150
840 ValGlyLeuLeuSerGlnGlyLeuArgIleAlaProProGluAlaProVa 856
|||||
151 GTGGGCATCTTACCAAGGACGACGAGAAATGACCTCCCGAAGCCCTGC 200
856 lThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspLeuValSerL 873
|||||
201 AACTGGTTATATGTTGGCCCAAGGGATGTACTTTGCTGACCTTGTCAGCA 250
873 ysSerAlaGlnTyrCysTyrValAspArgAsnAsnProValGlyLeuMet 889
|||||
251 AGAGTGTCTAGTATGTTCTCACTGATAAGAAATCTGCTGCTGCTAATG 300
890 LeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLysLysAlaTh 906
|||||
301 CTTTGTAGTGAAGTTGCCCTTGGAAATGCTGATGAGCTCAAAAAGCTAA 350
906 rSerMetAspLysProProArgGlyLysHisSerThrLysGlyLeuGlyL 923
|||||
351 GTATATGGATAAACCCTCCGAAGGAAAGACATTTCTACTAAAGAGACTGGCA 400
923 ysThrValProLeuGlu 928
|||||
401 AGAAATGCCCCCTGGAA 417

seq_name: gb_est2:BG751755
seq_documentation_block:
LOCUS BG751755 771 bp mRNA EST 15-MAY-2001
DEFINITION 602730481F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4874012 5',
mRNA sequence.
ACCESSION BG751755
VERSION BG751755
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
1 (bases 1 to 771)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1753 row: a column: 21
High quality sequence stop: 758.
Location/Qualifiers
1. .771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4874012"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pGB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. I"

BASE COUNT 221 a 177 c 177 g 196 t
ORIGIN
alignment_scores: Quality: 586.00 Length: 232
Ratio: 3.256 Gaps: 5
Percent Similarity: 77.586 Percent Identity: 50.862
alignment_block:
US-09-236-995D-2 x BG751755 ..
Align seg 1/1 to: BG751755 from: 1 to: 771

762 LysTyrMetLysLeuHisCysAsp..... 769
|||||
50 AAGGCCAGACACCCATTTGACCAACACATATAGTAGCTCCTCAGATAGA 99
770lThrProLeuAlaHisAspSerGluAspTyrLysLeuLeuGluG 785
|||||
100 GCACCTGGCCGCCCTTGACCATGAAGTTATGAGTTCAAAGTGAATCCCC 149
785 lntTyrLeuLeuAsnThrHisAlaProThrHisLysAspTyrSerLeuGlu 801
|||||
150 AGTACCTACATCTACCATGCTCCACACACAGCAGCTATACCATGACC 199
802 LeuGluGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTyrSe 818
|||||
200 TTGCTGATTTGTTGAAGTGGAGAGGATGTTGAG.....AAAGA 240
818 rArgTyrLysAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlySerA 835
|||||
241 AGCCTTCAGAGAGGACCTTCATACAGAGGATGCTTATGGCATGTTCCA 290
835 rGluThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAlaPro 851
|||||
291 GGATGAGTAACCTGGGTGGGAATCTTGAGCCATGGGCTTCGAATGCCCA 340
852 ProGluAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheAl 868
|||||
341 CCTGAAGCTCCCATCACAGGTTACATGTTTGGGAAAGGAATCTACTTTC 390


```

770 leThrProLeuAlaHisAspSerGluAspTyrLysLeuLeuGluGlnTyr 786
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
109 TTAAGTGGTTGACAGAGATTCTGACAGCGCGGAGGTCATCAGCAAGTAC 158

787 LeuLeuAsnThrHisAlaProThrHisLysAspTrpSerLeuGluLeuG1 803
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
159 GTGAAGAACAACACTCATGCTACCAACGCGCTATGACCTGGAAGTGTAT 208

803 uGluValPheSerLeuAspArgAspGlyGluLeuAsnLysTyrSerArgT 820
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 CGATATCTTCAAGATAGAGCGCGGAGGCGGAGAGCCAGCGCTACAGCCCT 258

820 yrlYsAsnAsnLeuHisAsnLysMetLeuLeuTrpHisGlySerArgLeu 836
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
259 TCAGG...CAGCTTCACAAACCGGAG.CTGTCTGTGCACGCGCTCCAGGACC 304

837 ThrAsnPheValGlyIleLeuSerGlnGlyLeuArgIleAlaProG1 853
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
305 ACCAACTTTGCTGCATCCTCTGCGAGGGTCTGCGGATAGCCCACTTGA 354

853 uAlaProValThrGlyTyrMetPheGlyLysGlyLeuTyrPheAlaAspL 870
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 AGCGCTGTGACAGGCTACATGTTGGGAAAGGATCTACTTTGCCGACA 404

870 euValSerLysSerAlaGlnTyrCysTyrValAspArgAsnAsnProVal 886
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
405 TGGTCTCCAAAGTCAAACTACTGCCACACATCTCAGGGAGACCCGATT 454

887 GlyLeuMetLeuLeuSerGluValAlaLeuGlyAspMetTyrGluLeuLy 903
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
455 GGCTTTAATACTGCTGGGAGAGTTCCTTGGAAACATGTATGAACTCAA 504

903 slyAlaThrSerMetAspLysProArgGlyLysHisSerThrLysG 920
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 GCATGCTTCACATATACAGCAAGTTACCCCAAGGCAAGCACAGTGTCAAAG 554

920 lYleuGlyLysThrValProLeuGluSerGluPheValLysTrpArg 935
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
555 GTTTGGGAAAAAACCCACCCCTGACTCTTCGGCCAGCATCATCCCTGGAGG 601

```

```

seq_name: gb_est1:AA212857
seq_documentation_block: 617 bp mRNA EST 18-FEB-1997
LOCUS AA212857 mouse NML Mus musculus cDNA clone IMAGE:677377 5'
DEFINITION mw44d01.r1 Soares mouse mRNA for poly (MOUSE), mRNA sequence.
similar to gb:X14206 Mouse mRNA for poly (MOUSE), mRNA sequence.
ACCESSION AA212857
VERSION AA212857.1 GI:1811476
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 617)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marla M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:417081
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 489.
Location/Qualifiers
1..617
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:677377"
/tissue_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAATCGGAGCGCGCGGATCTTTTCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 168 a 160 c 166 g 123 t
ORIGIN

```

```

FEATURES
Source
1..617
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:677377"
/tissue_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAATCGGAGCGCGCGGATCTTTTCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

```

alignment_scores:
Quality: 577.50 Length: 199
Ratio: 3.417 Gaps: 2
Percent Similarity: 84.925 Percent Identity: 56.281

```

```

alignment_block:
US-09-236-995D-2 x AA212857 ..
Align seg 1/1 to: AA212857 from: 1 to: 617

```

```

739 LeuGlnAspIleGluIleAlaSerLysIleVal.....GlyPheAspSe 753
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : :
9 CTCCTGGACATCAGGTCGCTATGCTCTCTCAGGGGTGGCTATGACGA 58

753 rAspSerAspGluSerLeuAspLysTyrMetLysLeuHisCysAspI 770
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
59 CAGCAGCAGGATCCCATCGAGTCACTACGAGAACTCAAACTGACA 108

```